2011

DEVELOPING A GENOTYPE-PHENOTYPE TABLE ONTOLOGY

Shifta Ansari

DEVELOPING A GENOTYPE-PHENOTYPE TABLE ONTOLOGY

(Spine title: Developing a Genotype-Phenotype Table Ontology)

(Thesis format: Monograph)

by

Shifta Ansari

Graduate Program in Computer Science

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

THE UNIVERSITY OF WESTERN ONTARIO

School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor:

.....

Dr. R. E. Mercer

Examiners:

.....

Dr. J. L. Barron

.....

Dr. S. Osborn

.....

Dr. G. Denham

The thesis by

Shifta Ansari

entitled:

Developing a Genotype-Phenotype Table Ontology

is accepted in partial fulfillment of the

requirements for the degree of

Master of Science

.....

Date

.....

Chair of the Thesis Examination Board

Abstract

Biomedical researchers show their experimental findings in a tabular form, which is one of the most important information sources in a scientific paper. The aim of this thesis is to develop and implement an intelligent tool that extracts and interprets the information from phenotype-genotype tables found in scholarly biomedical papers. In order to extract the information effectively, a table-based ontology to describe the relations among fields in these tables was developed. The ontology is domain-based. This thesis concentrates on the biomedical papers specially selected by the keywords mutation, gene, phenotype, genotype, disease, and syndrome. The tool is composed of an automated system to extract information from tables and to store the information in the ontology. The tool is verified by populating the ontology with some table information taken from the set of tables from which the ontology is designed. The tool is also verified by populating the ontology with a new set of tables to see what problems still exist for continuation of this work.

Keywords: table interpretation, ontology, information extraction.

Acknowlegements

First of all, I would like to thank Almighty Allah for everything. Without You I will be nothing, without You I am nothing.

Behind my thesis work, there are many peoples' contribution. I am now taking pleasure by thanking them.

It is my pleasure to thank my supervisor Dr. Robert Mercer to make this thesis possible. He is not only a great source of knowledge but also his continuous help was a great support for me. Whenever I was in trouble finding solution for my problems his door was always open for me. Dr. Mercer is the best supervisor that anyone could ever wish for. I enjoyed working with him for the past one year and eight months.

I would also like to thank Dr. Peter Rogan for his valuable time that he gave me to clear my concept about the biomedical terms while building up the ontology. Furthermore, his useful advise and opinions was really helpful for the thesis.

I would like to thank my colleague and friend Maryam Khordad for supporting me time to time. If I got worried about something, her words "Do not worry" or "It will be fine" always cheered me up.

I am extremely indebted to Rushdi Shams who was always there for me in my darkest hours. Whenever I was in trouble I always find him beside me and he gives his best effort as if this is his problem. This can be expected only from a true friend.

Thanks to Quazi Mah-Zereen Akter for counting me as a family member. Thank you for supporting when I got frustrated for some reasons and needs some mental supports. She is always there for me.

I would like to thank Art Bugorski for helping me with my problems even in his busy days.

I would also like to thank Janice Wiersma for making everything as convenient as possible.

I am extremely indebted to my father and mother for all the supports for making my dreams come true. Word cannot express how lucky I feel to be your daughter. All

good things that happens to me only because of your prayers and good deeds over the years. May Allah give you all the happiness of the heaven in earth and hereafter.
Thanks to my mother for supporting me from half a globe away while I was worried about my thesis. Thanks to my father for standing and praying for me every night. I consider myself as the luckiest person in this world to have you as my parents.
Thank you my brother for being a pillar of support. Thanks for all the prayers and advices. I am lucky to have you as my brother.
I would like to thank my uncle for all the good advices he gave me for my studies and for the prayers as well. You are the ideal of my career.
Thank you my sisters for giving me your valuable time listening to my problems over the phone and supporting me with the good words.
May Allah bless you all.
To the members of my thesis committee: John Barron, Sylvia Osborn, Graham Denham and Jamie Andrews. I appreciate your serving on my committee and the excellent feedback provided.
I am grateful to Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant for funding my work.
Finally, I would also like to thank the Department of Computer Science for providing me the funding and opportunities to make my dreams come true.
This thesis is dedicated to the memory of Rahnuma Afrin.

Contents

List of Figures

Chapter 1

Introduction

Scholarly writing in the broad area of experimental biomedicine is a genre that has a rhetorical style that exhibits some easily identifiable stylistic features: division of the paper into well-defined sections (Introduction, Methods, Results, Discussion), and the use of tables and figures to organize and express important results. Tables and figures have stylistic features, as well: titles, captions, content.

In addition to being easily detectable and having a reasonably standard form, tables are one of the most important sources of information, especially in scholarly biomedical papers. The reason for this is the community-accepted rhetorical style for authors of scientific papers to publish their experimental findings in a tabular form. This use of tables to report experimental findings happens in biomedical papers because the quantity of experimental data is large and the tabular arrangement allows for a concise presentation of the relationships among the data and for a better understanding of the results.

My interest in automating the extraction of data from tables in biomedical papers is motivated by the rapid advancement of knowledge in the biomedical field. The rate of publication of biomedical papers (more than 1800 in 2005 [15] and increasing) makes it impossible for researchers to stay abreast of the new findings. This situation has led to significant efforts to build automated tools to extract information from a paper and

interpret it automatically. Current NLP (Natural Language Processing) systems try to extract from the biomedical literature different knowledge such as, protein–protein interactions [8] [10] [17] [18] [32], new hypotheses [12] [13] [27], relations between drugs, genes and cells [7] [25] [28], protein structure [9] [14] and protein function [2] [29]. Most of this effort has been directed toward information extraction from text, although some recent attempts have been made to extract information from figures [4] and tables [30].

My contribution to this effort is a general approach to understand the information contained in tables and to extract and make it available for further use. Because the scope of table understanding is so great, I am beginning this undertaking by concentrating on one particular subclass of tables in the biomedical literature: tables that report relationships between phenotypes and genotypes. This research integrates with other projects: the analysis of phenotype-genotype relationships in the paper text and the use of this information in a biochemical tool. The field of biomedicine is made up of different disparate subdomains. I am mainly interested in genetics more specifically phenotype, genotype, mutation, and gene, and their relationships: syndrome (constellation or spectrum of phenotypes), and disease.

To extract the table portion from a paper accurately I take into consideration some factors like in what format (XML, PDF, HTML) usually authors like to write their paper. One study shows that the most popular format among the publishers of e-books is PDF [20]. According to their study, 8 out of 9 e-book publishers prefer PDF format. Springer e-book publishers provides both PDF and XML format. On the other hand Taylor and Francis publish their content in both PDF and HTML format. Questia prefers XML over PDF or HTML format. However, some publishers use PDF format and HTML format together such as Knovel and Gale Virtual Reference Library. Blackwell Reference Library is the only online e-book publishers which uses HTML format.

In the case of e-journals the scenario is similar. The most preferable format of

publication is PDF and/or HTML format (from Wiki). A small number of publications are in DOC format. However, there is a growing trend of using the XML format among journal publishers.

In my thesis I take into consideration PDF format of the paper. I used the table extraction software *TableSeer* to isolate the table from the pdf file. In some cases I was not successful extracting a table. In these cases I used *pdf2html* software which also gives the ability to provide xml output. From this XML output I isolated table portions along with the footnotes.

To summarize my contribution: I have collected and used papers from this domain to engineer a tool which populates an ontology with data reported in a table. I employ the concept of reading path [16], modifying its original specification to handle certain special cases like row oriented tables. Furthermore, one of the most important things is to retain the relationship between columns and/or rows of the tabular data so that further query of the ontology can be informative. I maintain the relation among the data of the table to the ontology. At the end, I search the ontology to find out particular information like mutation, corresponding gene where mutation occurs, position of the mutation, phenotype, genotype, etc.

The rest of the thesis will be organized as follows: Chapter 2 will provide a Literature Review of the Table isolation and Table interpretation research works which provide the background knowledge for the thesis. This chapter also includes some definitions and examples of some important terms along with some ideas that are considered as a foundation of the thesis. This will be followed by the methodology of the work of the entire thesis which will be described in Chapter 3. The methodology of entire thesis work is divided into categories like Table Isolation Technique, Table Information extraction process, description of the ontology and how the ontology is populated. Chapter 4 includes examples of populating the ontology for different tables, the problems while populating the ontology and the solutions of the problems. Chapter 5 will conclude the work presented, summarizing the work undertaken, and

will discuss possible future works for research on this thesis. Appendix A, B, C, D contains the results of some queries. Appendix E contains the references of all the papers that I have curated to design the software.

Chapter 2

Literature Review

This chapter provides a survey of background information needed to understand the process of table information extraction which consists of the isolation of the table text in the paper followed by interpretation of the information contained in the table. This chapter will also describe different ontological views of the table by Matthew Hurst [16] along with the reading path calculation which can be defined by reading of a table from the sequence of the reading of cells.

The different table isolation techniques and table understanding factors will be described in Section 2.1. This will be followed by a discussion of different ontological views of the table in Section 2.3. These four ontological views are: the physical (Section 2.3.1), the functional (Section 2.3.2), the structural (Section 2.3.3), and the semantic (Section 2.3.4). After these views are described, different representations of the table can be given. These representations are the physical, functional, structural representation of the table will be described in Section 2.4. Then, how table reading can be committed by reading of the sequence of cells is described in Section 2.5. Finally, Section 2.6 will describe the *Unified Medical Language System (UMLS)* ontology.

2.1 Table isolation and interpretation techniques

Isolating tables in documents has been an important research problem. Many researchers have introduced and implemented table isolation algorithms to allow tables to be lifted from the surrounding text in a variety of media. Each algorithm has some minor flaws. Some of the works that were incorporated in this thesis are discussed below. Electronically accessible archival biomedical literature is available in PDF and XML formats. Because the isolation of tables in XML-formatted literature is straightforward (the tables are bounded by the XML <table> </table> tags) the discussion below is only concerned with isolating tables in the PDF format.

2.1.1 *TableSeer*: Table isolation software by Liu

Y. Liu [19] developed the open-source table isolation software *TableSeer* to isolate tables from papers in the PDF format. This software applies the page box cutting method and a heuristic sparse-line detection technique for table edge detection. A page box is defined as a rectangle consisting of adjacently connected lines, that is, they have a uniform font size and they are in the same document page. The heuristic rules map boxes to different logical components (titles, authors, affiliations, abstract, references, etc.) and specific physical components (tables, figures, etc.). Tables are considered to be sparse, whereas document text is not sparse. Not only is the table content extracted with all of the important cell information used to create T^{phys} (a suitable representation of table information defined by Mathew Hurst [16], but also the table title, caption, and footnotes. Furthermore, after isolating the table information from the PDF file, the software has its own table ranking algorithm to rank the table. One of *TableSeer*'s failings is the case of multi-line footnotes. Only the first line of multi-line footnotes belong to a table.

2.1.2 *pdf2table*: Table extraction software by Yildiz

B. Yildiz [31] built the *pdf2table* system on top of *pdf2html* developed by G. Ovtcharov and R. Dorsch [22]. *pdf2table* isolates tables from PDF files by converting PDF files into strings with the corresponding coordinate values in the PDF file. This information is saved in the XML file format for further table detection processing. Yildiz used the concepts of multi-line blocks and single line blocks to store data. The different blocks allow multi-line footnotes to be processed correctly. However, *pdf2table* correctly processes only single column pdf files, whereas many PDF files are multi-column PDF files.

2.1.3 *Tab Pro*: Table isolation and interpretation software by Hurst

In M. Hurst's thesis [16], undoubtedly the most comprehensive work on automated table understanding, contains a thorough and detailed discussion of table interpretation factors. He analyzed table-containing documents from different sources: news articles taken from paper sources; documents scanned and converted into text files using OCR software; LaTeX files from the Computational Linguistics e-print archive; HTML files that facilitated easy table extraction by searching for the <table> tag. The LaTeX and HTML files where further processed by converting them into system DTD whereas documents from other sources were marked up by hand.

In addition, he developed and integrated some basic rules to represent generic table information. He developed the concept of *reading path* for cells, building category of the data, building physical, functional, structural, and semantic ontologies of the tables to interpret table information and if any ambiguity occurs then the correlation between ontologies can be helpful to remove the ambiguity. Furthermore, he developed the concepts of physical and functional models of the table which are created through the combination of the four ontological elements. The reading path concept

is derived from the functional model of the table and the simple table relation concept. These all facilitate the interpretation of a table. Each of these concepts will be discussed in detail in Section 2.3, but first, I turn to defining some key terminology in the next section.

2.2 Preliminary Definitions

A table can be subdivided into three components: the *caption*, the *footnotes*, and the *table content*. Normally, the table caption occurs before the table content. The caption begins with the keyword "Table", followed by a tag (a number, typically) to allow tables to be identified, followed by some text that describes the content of the table. The footnotes, if they exist, follow the table content. This text is often used to describe the symbols and abbreviations that occur in the table content. The table content is the remaining information. The table content is organized into *rows* and *columns*. A table *cell* occurs at the intersection of a table row and a table column. Hurst [16] divides all the cells of the table into two functional categories: *access cells* and *data cells*.

In a table, a cell which is characterized as terminating, that is data appears in a terminating role, is called a *data cell*. Considering the table in Figure 2.1, the yellow blocks are the data cells. All of the remaining cells are the *access cells*. These cells can be conceptualized as indices or descriptors of the data cells. In the table in Figure 2.1 the blue blocks are the access cells.

Depending on which aspect of the table (row or column) is to be considered the main axis of the information, tables can be given an orientation. Based on the orientation of the data in the table, tables can be categorized as row-oriented and column-oriented tables. I will refer to the first type of table as a *horizontal table*, and the latter type as a *vertical table*. The table orientation affects the order that access cells are examined. It is not a simple task to determine what is to be the

table orientation, so for the purposes of this thesis, I have manually interpreted the orientation of the table.

2.3 Ontological Views of Tables

Locating tables in various printed and electronic media and understanding the content of these tables has been the focus of a number of research projects. A few of these projects were discussed in the previous section. The most comprehensive project is the one presented in the PhD thesis of M. Hurst [16]. In this and the next section I will summarize the aspects of his thesis that are important for my work. In the ontological description of the table, Hurst describes various ontological views of the table: the physical, functional, structural, and semantic.

2.3.1 Ontology Description: Physical

Starting with the physical ontological description of the table, Hurst describes the physical table using the relative position of the cells. He considered this as a basic informational element. To calculate the position of the cell (relatively) he imposed a minimal grid over the table and considered labeling the rows and columns of the table with natural numbers. Then each cell is defined physically by the value of the grid in top-left and bottom right corners. For the example in Figure 2.1, the table is first divided into the minimum possible sectors called grids and then the columns and rows are numbered individually. Now each cell can be defined physically by calling out the numbers of top-left and bottom right corners of the cell. The cell containing the data "Number of Students" can be physically described in terms of cell numbers as [(1,0),(2,0)]. Similarly "Business" can be defined as[(0,3),(0,3)].

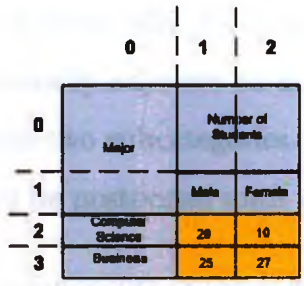

Figure 2.1: Dummy Table

2.3.2 Ontology Description: Functional

According to Hurst [16] a functional view of the table concerns two issues, such as distinguishing cells according to their function and reading information from the table.

Cells can be divided into two categories: access cells and, data cells. Data cells are always the targets of the search for information in the table, whereas access cells define the path to get a data cell to facilitate better understanding of that data cell in the table.

Reading of a particular table consists of reading to data cells in the table. There must be a way to get to the data cell. Reading a table is a combination of the navigating cells, in using ordering information.

2.3.3 Ontology Description: Structural

By Structure, Hurst means the aspect of the table with a restriction about navigation of the cells by a local search operation. Structural description of ontology also employs the concept of Simple Table Relation: a structural representation from which logically adjacent cells can be determined from a particular position in the table. Cell organization can be demonstrated in two ways:

1. Reading path: To read particular information from the table, the concept of a reading path can be used. This path indicates how to reach a particular

cell through a sequence of other cells. Cells in a sequence of cells are logically adjacent rather than physically adjacent to each other. The reading path can be further categorized into two subcategories. Further discussion of these two types of reading path will be postponed until Chapter 3.

2. Canonical Table: If the components of the table reduce to a relation similar to a relational database, then the converted table form is known as a canonical table. The canonical table preserves the restricted connection of cells without the structural information, such as the order of the cell, category organization, etc.

   Structure facilitates the physical economy of the table. For example, cells containing the same value can be grouped together to form a single cell to save spaces to store the whole table. Moreover, similar information can be grouped together to imply a certain concept, like placing some statistical information for different years together so that the reader can understand and compare conveniently. This can be facilitated by organizing the cells.

2.3.4 Ontology Description: Semantic

While interpreting the table information, some problems can arise like missing linguistic information (e.g. N/A in any data cell may mean that information is not available). "Patient" and "Patients" column headings may indicate the same object, or "Reference" and "Reported case" may convey the same meaning between the cell contents on a reading path. According to Hurst, to solve these problems a number of assumptions can be made which will help to come up with a linguistic semantic model of the table. Some of the assumptions are as follows:

1. Interpretation of cell contents: Hurst describes the interpretation of cell contents at the object level and at the meta level. At the object level, he describes how a cell of the table can be described in conjunction with other cells in the

table linguistically and pragmatically. In this thesis I need to determine the column header at various linguistic levels e.g. "Locus" and "Gene Position" (domain knowledge), "Patient" and "Patients" (morphology), "Reference" and "Reported Case" (synonyms in this domain) each refer to the same thing.

Moreover, at the object level, sometimes pragmatic analysis of the term is needed For example, in Figure 2.2 in the first column certain terms need to be defined to populate the ontology. I use external sources like MetaMap and OBO-Edit to determine what ontological concept these terms map to. This is further explained in Chapter 3.

Besides this, at the meta level Hurst describes what to expect in the cells depending on what the cell represents. However, I did not apply this concept in the thesis.

2. Intercell relationship:

   According to Hurst, an intercell relationship exists between the interpretations of cell contents. Hurst suggests that if cells are found next to each other, then it is possible to apply natural language-like semantic interpretation with those adjacent cells. (He used the terminology 'sentence level' instead of 'natural language-like'.) He further describes that this relationship can also be defined by interpretation of the discourse that can be found in the document as a whole. In my approach I have used both methods. I follow the prior method, for example, when I adjoin the category headings "Brain anomalies" with the subcategory heading "Corpus callosum" (Figure 2.2). The latter method is followed when certain unknown terms appear in the column heading in a table. To analyse the term, further extra-table information need to be retrieved from other portions of the paper.

   Besides these assumptions for extracting linguistic relationships from the table, Hurst also concentrated on ontological relationships. For example, Nominal

Table 3: Clinical features and size of deletion of the 12 patients with 13q monosomy.

Patients	1	2	3	4	5	6	7	8	9
Deleted segment	13q13.3-13qter	13q21.1-13q31.1	13q21.32-13qter	13q31.1-13q33.3	13q31.1-13qter	13q31.1-13qter	13q31.3-13q31.1	13q31.3-13q34	13q32.1-13qter
Size of deletion	70 mb	30 mb	47 mb	28 mb	34 mb	30 mb	10 mb	20 mb	18 mb
Sex	f	m	m	f	f	m	m	m	m
Child(c)/foetus(f)	f(33wg)	c(13m)	f(25wg)	f(24wg)	f(25wg)	f(26wg)	f(32wg)	f(23wg)	f(21wg)
IUGR	+	NK	+	-	+	+	-	-	+
Growth retardation	NK	+	NK	NK	NK	NK	NK	NK	NK
Microcephaly	+	-	-	-	+	+	-	+	+
Mental retardation	NK	+	NK	NK	NK	NK	NK	NK	NK
Brain anomalies									
Corpus callosum agenesis	-	-	+	-	+	NK	NK		
Holoprosencephaly	-	-	-	-	+	+	+	+	
Cerebellar vermis hypoplasia	+	-	+	-	-	NK	NK		
Digestive anomalies									
Common mesentary	-	-	-	+	-	-	-	-	-
Pancreas anomalies	-	-	-	+	-	-	-	+	-
Gall-bladder agencies	-	-	-	-	-	+	-	Hypoplasia	-
Abnormal lungs	+	NK	+	+	+	+	-	-	+

NK: not known, m: months, y: years, wg: weeks of gestation, f: female, m: male; arsa: aberrant right subclavian artery; dscv: double superior vena cava. (1) nasal bone hypoplasia.

Figure 2.2: Example of a Table from a set of tables based on which I engineered the proposed ontology. This table is reproduced from [24] with some rows and columns removed to fit the page. The rows are rearranged for addressing different types of problems.

Super-type is [Car, Ford: Ford is one type of Car], or Partitive is [Car, Wheel: Wheel is a part of Car], or Quantitative is [No. of cars, 2]. I am working with a specific domain: genetics. More precisely, I am dealing with a subdomain containing concepts such as phenotype, genotype, mutation, syndrome, disease, and genes. I followed a similar relationship extraction procedure based on my observation of the ontological aspects found in the tables curated from the selected domain.

2.4 Various representations of the table model

In Hurst's thesis [16] a number of representations of the table model are described. The representations which are important for this thesis are discussed below.

2.4.1 Physical representation

According to Hurst, the table is represented as a set of cells. A cell has a top left and a bottom right relative coordinate value and the contents of the cells are considered as strings. His definition of a cell is as follows:

> A cell, C, is a 6-tuple $(id, x^1, y^1, x^2, y^2, string_0)$ where id is the cell identifier ($id \in T$), x^1 and y^1 are the upper-left coordinates, x^2 and y^2 are the lower right coordinates and the *string* is the cell content.

Physically, a table is defined in terms of cells. The physical table, T^{phys}, is a set of cells described in the above manner: $\{(id_0, x_0^1, y_0^1, x_0^2, y_0^2, string_0) \ldots (id_n, x_n^1, y_n^1, x_n^2, y_n^2, string_n)\}$ with the following two constraints:

- no cells intersect:

$$\forall C \in T^{phys}, \neg C' \in T^{phys} : (C = C') \vee (((C'.x_1 \leq C.x_2)) \wedge ((C'.x_1 \geq C.x_1)) \vee ((C'.y_2 \leq C.y_2) \wedge (C'.y_2 \geq C.y_1))) \wedge$$
$$(((C'.x_1 \leq C.x_2)) \wedge ((C'.x_1 \geq C.x_1)) \vee ((C'.y_1 \leq C.y_2) \wedge (C'.y_1 \geq C.y_1))) \wedge$$
$$(((C'.x_2 \geq C.x_1)) \wedge ((C'.x_2 \leq C.x_2)) \vee ((C'.y_1 \leq C.y_2) \wedge (C'.y_1 \geq C.y_1))) \wedge$$
$$(((C'.x_2 \geq C.x_1)) \wedge ((C'.x_2 \leq C.x_2)) \vee ((C'.y_2 \leq C.y_1) \wedge (C'.y_2 \geq C.y_2))).$$

- no strings are empty:

$$\neg \exists C \in T^{phys} : C.string = \text{“”}$$

2.4.2 Functional representation

The functional table, T^{func}, is a tuple (A, D), where A is the set of access cell identifiers and D is the set of data cell identifiers.

1. $\forall X \in A, X \in T.C$
2. $\forall X \in D, X \in T.C$
3. $A \cup D = T.C$
4. cell X is an *ACCESS* cell if $X \in A$
5. cell X is a *DATA* cell if $X \in D$

Normally $A \bigcap D = \phi$

2.4.3 Simple Table relation

The simple table relation (STR) is a set of triples, $\langle X, Y, R \rangle$ where X is a cell identifier representing the source and Y is a cell identifier representing the sink of a directed arc and R a set of n restrictions (r_0 *to* r_n) on the transition of the arc: $(id_0, id_1, \{id_{r_0}, \ldots, id_{r_n}\})$. Since none of the tables that I deal with have restrictions, I will omit this from all of the discussion below, although Hurst's definitions contain this concept.

The structural table, T^{struc}, is, then, described as a set thus:

$$\{(id_0^0, id_0^1), \ldots, (id_n^0, id_n^1)\}$$

with the following constraints:

- the relation holds between two different cells:

  $\forall C \in T^{struc} : \neg C.X = C.Y$ and

- any pair of cells can appear only once in the relation:

$$\forall C \in T^{struc}, \neg\exists C' \in T^{struc} : C.X = C'.Y \wedge C.Y = C'.X$$

2.5 Reading of a table

From Hurst [16] the reading of a table is the set of readings for all data cells. The reading of a particular data cell can be defined by the following equation [16]:

- $\vec{X}$ is the set of paths to X.

The set of cells P, a path to cell X, is defined recursively:

$$\exists\, y \in T : \langle Y, X\rangle \epsilon\, T^{struc}, \exists P' \in \vec{Y}, \textit{ then } P = \{Y\} \cup P'.$$

$$\neg\, \exists\, Y \in T : \langle Y, X\rangle \in T^{struc}, \textit{ then } P = \{\}. \tag{2.1}$$

a reading of a cell is the set of paths to that cell. The following additional constraint is required.

$$\forall P \epsilon \vec{X},\ P \textit{ appears exactly once} \tag{2.2}$$

So, to read the data from the whole table, first all the data cells should be read. To read a particular data cell, all the paths to that cell should be determined.

2.6 Unified Medical Language System ontology

The biomedical research community has developed extensive systems and ontologies to organize information and supply tools. One such system is the *Unified Medical Language System (UMLS).* UMLS contains information linking language and knowledge in the biomedical domain. Natural language systems can use UMLS to help parse sentences found in the biomedical literature. Sophisticated information retrieval systems can be then be designed around these natural language systems. For different applications UMLS will have to be customized to purpose one's needs.

Metathesaurus is the base of the UMLS and is comprised of over 1 million biomedical concepts. Moreover, it includes 5 million concept names. Metathesaurus is organized by concepts. Each concept is assigned one or more semantic types (categories). These categories are linked through semantic relationships. The semantic relationships between concepts can be represented by a semantic network. In Figure 2.3, a portion of the UMLS Semantic Network and Relations is given.

In the network the semantic types are referred to as nodes. The relationships between them are expressed as links. The major groupings of this semantic networks are ORGANISM (maps the animal and plant kingdom), ANATOMICAL STRUCTURE (maps structures and related abnormalities, as well), BIOLOGICAL FUNCTION, CHEMICALS, EVENTS, PHYSICAL OBJECTS, and different concepts or Ideas.

ORGANISM contains different types of elements including plants and animals. Plant is directly connected with organism with an IS-A relationship whereas animal is connected through the node EUKARYOTE. ANIMAL has one child node and five grandchild nodes to cover all members of vertebrate animals such as AMPHIBIAN, BIRD, FISH, MAMMAL and REPTILE. The MAMMAL concept has HUMAN as a child node. They all are connected with IS-A relationships. Moreover, the subtrees under BIOLOGICAL FUNCTION are PHYSIOLOGIC FUNCTION and PATHOLOGICAL FUNCTION. PHYSIOLOGIC FUNCTION contains child nodes to describe functions of different components such as ORGANISM, ORGAN, CELL, and MOLECULAR FUNCTION. CELL and MOLECULAR FUNCTION, have a sibling class in a different subtree but with common grandparents which actually describes the dysfunction of the CELL and MOLECULE in the single semantic types (node in the network). The parent of this node is PATHOLOGICAL FUNCTION. PATHOLOGICAL FUNCTION has two child nodes other than the CELL or MOLECULAR DYSFUNCTION. These two nodes represents two concepts DISEASE OR SYNDROME and EXPERIMENTAL MODEL OF DISEASE both of which I am very much interested in. This DISEASE OR SYNDROME concept has only two children, MENTAL OR BEHAVIORAL DYSFUNCTION and NEOPLASTIC

PROCESS. Each nodes under the BIOLOGICAL FUNCTION are linked to its parent by the IS-A link.

Also, the semantic type ANATOMICAL STRUCTURE has three children EMBRYONIC STRUCTURE, ANATOMICAL ABNORMALITY and FULLY FORMED ANATOMICAL STRUCTURE, and each of these except EMBRYONIC STRUCTURE in turn has several children and grandchildren. ANATOMICAL STRUCTURE has two children CONGENITAL ABNORMALITY and ACQUIRED ABNORMALITY. Similarly, FULLY FORMED ANATOMICAL STRUCTURE has several children (BODY PART, ORGAN OR ORGAN COMPONENT, TISSUE, CELL, CELL COMPONENT, GENE OR GENOME) connected with it with IS-A relationships. However these children are connected with each other with relationships other than the IS-A relationship. For example, GENE OR GENOME is connected with CELL COMPONENT with the partof relationship. Similarly, CELL COMPONENT with CELL, CELL with TISSUE and TISSUE with BODY PART, ORGAN OR ORGAN COMPONENT all have partof relationship with each other. However, BODY PART, ORGAN OR ORGAN COMPONENT has a conceptual part of relationship with a concept BODY CONCEPT which does not have any relationship with the root FULLY FORMED ANATOMICAL STRUCTURE and any of the nodes of the subtree. There is another major concept named FINDING which connects ORGANISM Attribute with EVALUATION OF relationship. This ORGANISM ATTRIBUTE is connected with the major node Organism with purpose of relationship. This Finding has two children LABORATORY OR TEST RESULT and SIGN OR SYMPTOM. Each child in the hierarchy is linked to its parent by the IS-A link.

Similarly, Biological Function has a relationship with Organism which is PROCESS OF RELATIONSHIP. There are some other nodes which are not major semantic network nodes but still refer to some important connections in the semantic network. BODY SUBSTANCE is connected with FULLY FORMED ANATOMICAL STRUCTURE with the PROCEDURE CONTAINS relationship. Similarly, BODY LOCATION OR REGION and BODY SPACE OR JUNCTION is connected with FULLY FORMED ANATOM-

ICAL STRUCTURE with CONCEPTUAL PART OF RELATIONSHIP.

Injury or poisoning relationship is a single node in the network which has the DISRUPTS relationship with PHYSIOLOGICAL FUNCTION and FULLY FORMED ANATOMICAL STRUCTURE. The Injury or Poisoning node has recursive CO-OCCURS relationship with itself. Biological function has some connections with nodes other than the child nodes. It has PROCESS OF, EVALUATION OF, LOCATION OF relationships with ORGANISM, FINDING, BODY LOCATION OR REGION, respectively.

The proposed table ontology in the thesis has many similarities and differences with the UMLS ontology. The details are described in Chapter 3.

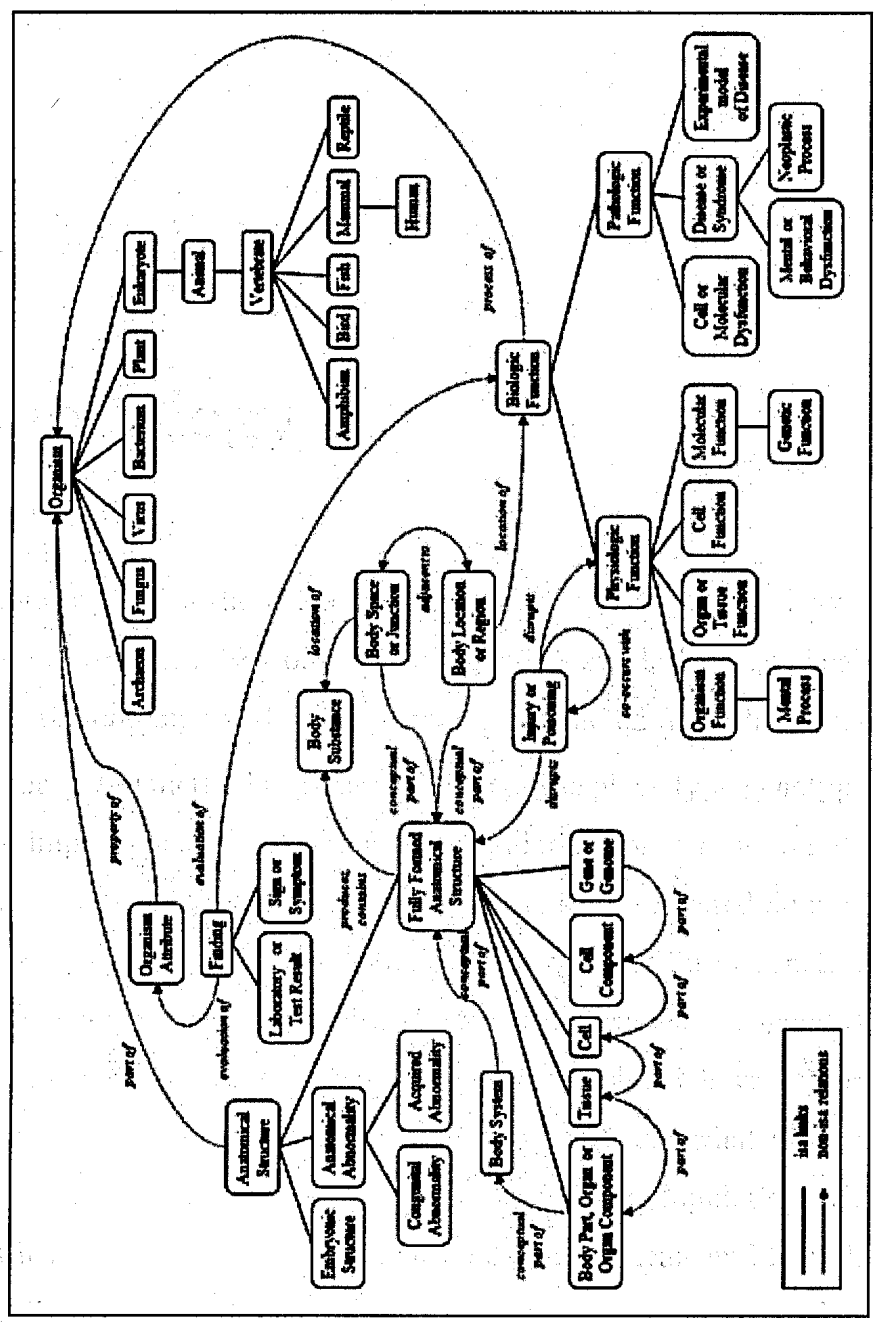

Figure 2.3: A Portion of the Unified Medical Language System (UMLS) Semantic Network, UMLS Terminology Services (UTS),
URL: https://uts.nlm.nih.gov/home.html

Chapter 3

Methodology

My contribution in this thesis is a phenotype-genotype table ontology and an automated method to take tables of this type found in scholarly biomedical articles and to populate the ontology with the information extracted from the tables.

In summary, the methodology used to design the phenotype-genotype table ontology and to implement the extraction and population software is as follows. To have a sufficient set of examples of this type of paper, I have curated 50 papers (listed in Appendix E) containing about 100 tables from the domain of genetics. More precisely, the papers collected deal with the subtopics *phenotype, genotype, mutation, gene, syndrome,* and *disease.* From 100 of these tables, I have designed the table ontology. Then I have collected a new set of papers to see what problems still exist for continuation of this work. In the next chapter, the population of the proposed ontology with some new set of tables, encountered problems and how the problem is solved are discussed. The table information extraction process can be divided into two main parts: isolating the table from the surrounding text, and extracting the information from the table and populating the ontology to provide a semantics for the information. All of the curated papers are in PDF. I have assembled and written some software that does this isolation semi-automatically, that is, I decide which software module to use to isolate each table. The table extraction method incorporates

Hurst's reading path concept that was discussed in Chapter 2 into a method that I have developed and implemented. The ontology is implemented using the Resource Description Framework (RDF).

Each of these contributions is discussed in detail below. Section 3.1 describes the development of the phenotype-genotype table ontology. Section 3.2 describes the software to isolate the tables. Section 3.4 demonstrates the ontology population software on a small set of examples.

3.1 The Phenotype-Genotype Table Ontology

The process of building a phenotype-genotype table ontology requires the curation of a sufficient number of texts containing a variety of tables in order to give credence to the proposed ontology. Using 50 papers curated from the selected domain, genetics (focusing on important terms like gene, mutation, phenotype, genotype, syndrome, and disease), I have engineered a table ontology that incorporates some new terms compared to the UMLS ontology to include relationships that exist in these tables (e.g. subjects have an age) as well as other fundamental concepts.

3.1.1 The Proposed Ontology

After investigating 100 tables from 50 biomedical papers, the semantic network, a portion of which is shown in Figure 3.2, has been proposed as a reasonable prototype. The semantic network has been designed based on the observed tables' access cells and the relationships that exist among the cells in the table. To organize the terms hierarchically, concepts need to be grouped. For example, CELL, DNA, GENE, CHROMOSOME seem to belong to the same group. To find an appropriate label for the root node of the concepts belonging to the same group and to partially verify the organization of these concepts, the UMLS semantic network is used. This methodology is appropriate for those terms that are in common between the proposed ontology

and the UMLS ontology.

An example of how certain decisions were made in the design is now given. I am experienced with new terms not associated with the keywords mentioned above but to facilitate interpretation of table information those terms have to be mapped to the ontology. For example, PATIENT has certain attributes like Age, Weight, Gender, Height. In the UMLS ontology, there is a concept named PATIENT or DISABLED GROUP which exists with the parent node GROUP. As well, there is an Age group under the concept GROUP. However, the information concerning PATIENT in the table represents the information about each patient rather than being presented as information about a group of patients. I have added a concept named ORGANISM to accommodate certain important classes like PATIENT and FAMILY and certain important attributes for them like age, gender, height, birth weight, etc. because the authors of the biomedical journals consider these features important when publishing their results in tabular form. This organism entity is also present in the UMLS ontology which contains EUKARYOTE → ANIMAL → VERTEBRATE → MAMMAL → HUMAN as a subtree. Also, based on my observation about the tables I put LABORATORY RESULTS with a different label EXPERIMENTAL FINDINGS under THING → FINDINGS whereas in UMLS it is put under conceptual entity → FINDINGS.

I have designed a phenotype-genotype table ontology that for those concepts and relations found in the UMLS ontology those concepts and relations are copied. For these concepts and relations I am certain of the veracity of the table ontology. In those instances in which the UMLS ontology does not contain the necessary concepts like genotype and phenotype in its structure, I used an expert's knowledge to assist us. In the proposed ontology (Figure 3.1) genotype and phenotype is considered as a conceptual entity. Some other conceptual entities like MUTATION, BIOCHEMICAL TISSUE DEFECT, CHROMOSOME DELETION, etc. are considered as different types of genotype and are placed under the GENOTYPE concept. Also, some cell values act as an identifier of the whole data (as in Figure 3.8 "Family" or "Patient No.").

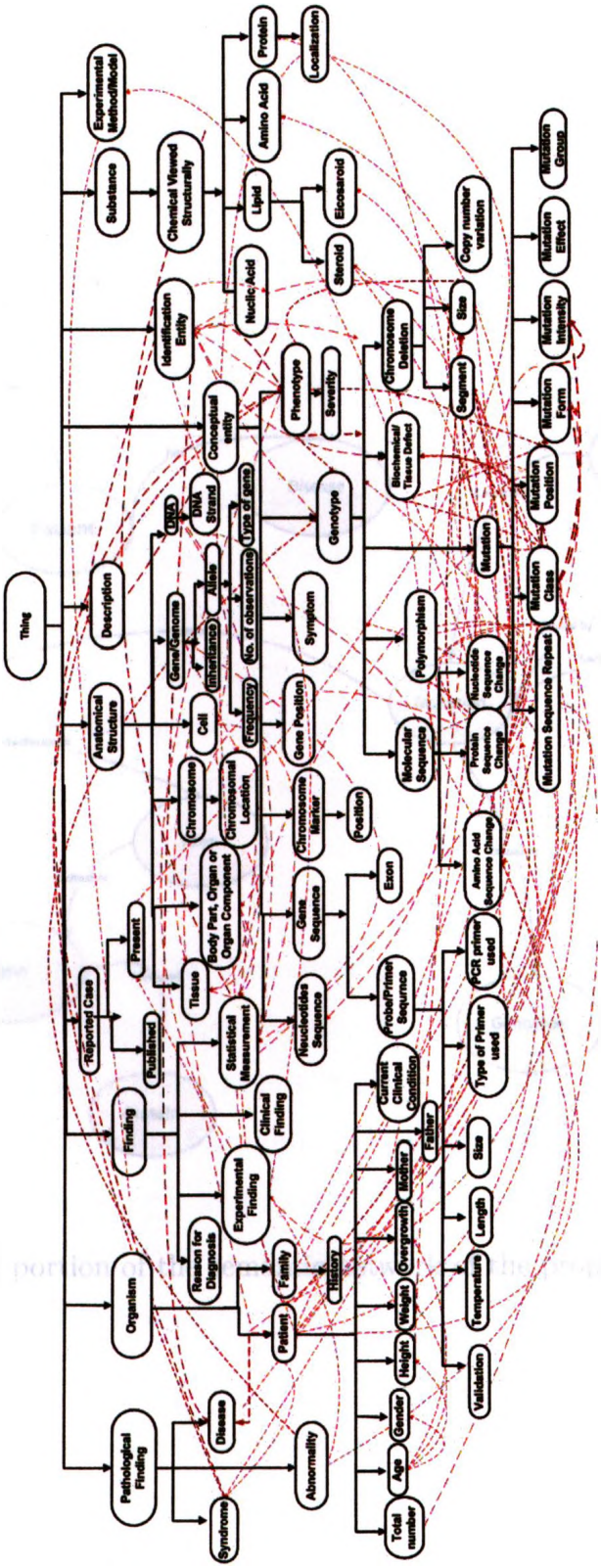

Figure 3.1: Ontology for the Observed Tables

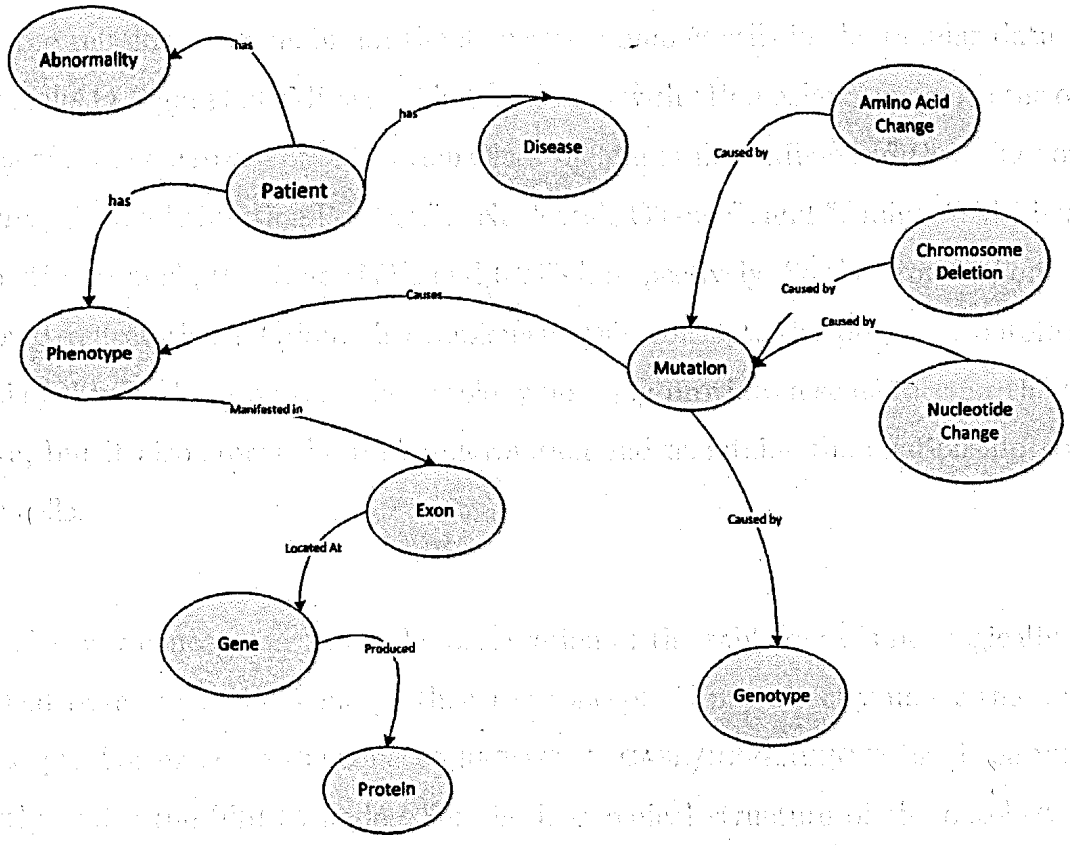

Figure 3.2: A portion of the semantic network of the proposed ontology

To accommodate this type of information in the appropriate place in the ontology I designed a concept named IDENTIFICATION ENTITY. In Figure 3.2 the partial semantic network of the proposed ontology is given.

It is necessary to accomodate not only the cell values but also the relationship among the cells that exist in the tabular form. In Figure 3.1 the red dashed lines are the relations that maintain the association among cells in the tabular data. For example, in Figure 3.8 "Missense" has a relation with "Homeobox" and with the other data of the same row. For "Mutation Class" holding value "Missense" the values of the corresponding "Mutation Position", "Nucleotide Change", and "Amino Acid Change" are "Homeobox", (CrG, nt 4171), and R247G, respectively. So the proposed ontology should reflect this relation. To implement this concept, the proposed ontology is designed in such a way that the ontology not only provides a semantics for the table data, but it also stores the table information and maintains the relationship among the cells.

If a new concept based on the observation of the table data is not logically subsumed by any existing concept then the concept, is put directly under the THING concept. For example the concept REPORTED CASE/REFERENCES has been put directly under the THING node. For the hierarchical structure of the ontology it is convenient to add a new concept under the root node. In the implementation phase I implemented it using RDF in the Java platform where adding a new concept with the necessary relationships is convenient.

The designed ontology should be interpreted as a prototype. Tables containing concepts not in the current implementation will be encountered. A semi-automatic method to determine this situation and to add the appropriate concept to the ontology is suggested as future work.

3.1.2 Building the Ontology

I implemented my proposed ontology using *Jena*, a Java framework for designing semantic web applications. More specifically, I used Resource Description Framework (RDF) of the Jena package to implement the ontology. It is an open source software. In RDF the concepts (e.g. MUTATION CLASS, MUTATION POSITION, MUTATION, GENOTYPE, PHENOTYPE) are implemented as resources. It is possible to create relationships between concepts and data as well. In Jena this feature is known as a property. By using properties, table information can be stored and the relationship among fields that holds between them can be maintained.

Considering the portion of the ontology in Figure 3.3, to create the hierarchical relationship between the concepts MUTATION and MUTATION CLASS, and MUTATION and MUTATION POSITION the sequences of code in java are as follows:

To create concepts MUTATION, MUTATION CLASS, and MUTATION POSITION the corresponding code:

```
Resource Mutation=m.createProperty( ontologyURI + "Mutation" );
Resource MutationClass=m.createProperty(ontologyURI+"MutationClass");
Resource MutationPos=m.createProperty( ontologyURI + "MutationPos" );
```

To create a hierarchical relationship between MUTATION and MUTATIONCLASS, I create a property called "MutMutClass" as follows:

```
Property MutMutClass = m.createProperty(ontologyURI + "MutMutClass" );
```

To connect two concepts or resources MUTATION and MUTATIONCLASS:

```
Mutation.addProperty(MutMutClass, MutationClass );
```

Similarly, to create a hierarchical relationship between MUTATION and MUTATIONPOS, I create a property called "MutMutPos" as follows:

```
Property MutMutPos = m.createProperty(ontologyURI + "MutMutPos" );
```

To connect two concepts or resources MUTATION and MUTATIONPOS:

```
Mutation.addProperty(MutMutPos, MutationPos );
```

Now to store data to concepts or resources, an additional property needs to be created. Here to store "Missense" value with the concept MUTATIONCLASS a property named "MutClassInd" is created and then added with the concept MUTATIONCLASS with value "Missense".

```
Property MutClassInd=m.createProperty(ontologyURI + "MutClassInd" );
MutationClass.addProperty(MutClassInd, "Missense");
```

Similarly here to store "Homeobox" value with the concept MUTATIONPOSITION a property named "MutPosInd" is created and then add with the concept MUTATIONPOS with value "'Homeobox". The codes are as follows:

```
Property MutPosInd=m.createProperty(ontologyURI + "MutPosInd" );
MutationPos.addProperty(MutPosInd, "Homeobox");
```

Now to retain the relationship between two cells "Homeobox" and "Missense" (that is for the MutationClass "Missense", the corresponding "MutationPosition" is "Homeobox) a property named MutClassMutPos is created which connect these two values "Missense" and "Homeobox" as follows:

```
Property MutClassMutPos=m.createProperty(ontologyURI+"MutClassMutPos");
MutationClass.addProperty(MutClassMutPos, "Homeobox", "Missense");
```

I can now move to the discussion of how information is extracted from tables that occur in PDF formatted papers. This discussion is split into three components: the isolation of the table itself in the PDF document in Section 3.2, the extraction of the information from the table in Section 3.3, and finally the population of the ontology with the extracted information in Section 3.4.

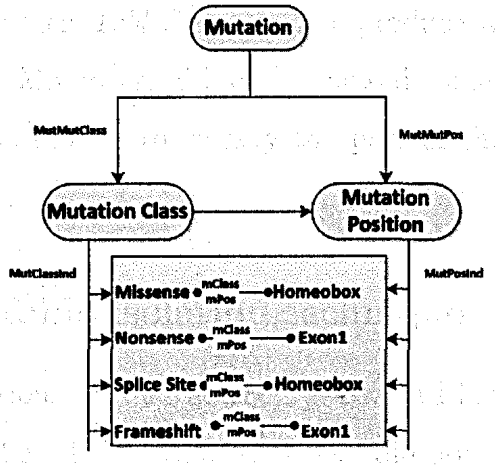

Figure 3.3: Details of a portion of the ontology

3.2 Table Isolation

The first step in the table information extraction process consists of isolating the raw, unprocessed table in the paper text followed by a transformation of the table's information into T^{phys}, Hurst's representation of the physical table. I detail the pre-existing methods that I used followed by the methods that I have developed to generate a robust extraction process. The presentation of a complete example of isolating and transforming the textual representation of a table is delayed to Chapter 4.

3.2.1 Table isolation open source software

To all of the tables contained in the 50 curated biomedical papers, I applied the *TableSeer* [19] open source software to isolate the tables from the text documents. It extracts the table portion along with the caption, footnotes, and the portion of the paper text that discusses the table information. This system applies page Box cutting method and a Heuristic-based Sparse-line Detection technique for table edge detection. It also has a system of extracting multiple tables at the same time in a batch mode, but I did not use this feature.

For some table instances *TableSeer* did not produce a successful extraction of multiple line footnotes. Moreover, TableSeer showed some difficulty recognizing all parts of the heading. It tended to classify the part of the heading near the table content as table content.

3.2.2 Modified Table isolation technique

To overcome the limitations of the TableSeer software I implemented some software to extract tables from the pdf files that TableSeer did not extract properly. In these cases I used *pdf2html* which also provides XML output. The XML format contains the top value, left value, width value, height value and the font of the text. See Figure 3.4 for an example.

To extract the table portion from the data along with the single or multiple line footnotes, I first implemented B. Yildiz's [31] multi-line block and single-line block concepts. Then I applied some heuristics to isolate the table portion. I detail these heuristics in the following section.

3.2.3 Heuristic approach for table isolation

A search for the keyword "Table" in the complete text is performed. Normally, this keyword appears at least twice, at least once in the body of the paper and once in the caption of the table. The heuristic to find the beginning of the table is based on the proximity of the keyword "Table" to a font value different from the font value of the body text of the paper. Furthermore, if the "Table" keyword is within the first 5 characters following the XML control information in the line containing the "Table" keyword then it is very likely that this is the caption of the table.

The end of the table is indicated by the font changing back to the body font. In this way the single line or multi-line table footnotes can be extracted. Normally, footnotes have a font value smaller than the font of the table content. So, once the table portion is isolated from the document, the footnote portion can be further

```
<text top="100" left="64" width="34" height="10" font="6">Table 1</text>
<text top="112" left="64" width="780" height="10" font="6">Clinical features
and size of deletion of the 12 patients
with 13q monosomy.</text>
<text top="125" left="64" width="394" height="10" font="6">right subclavian artery;
dscv: double superior vena cava. (1) nasal bone hypoplasia.</text>
<text top="147" left="64" width="36" height="10" font="6">Patients</text>
<text top="147" left="182" width="6" height="10" font="6">1</text>
<text top="147" left="236" width="6" height="10" font="6">2</text>
.....
.....
<text top="147" left="794" width="11" height="10" font="6">12</text>
<text top="159" left="64" width="77" height="10" font="6">Deleted segment</text>
<text top="159" left="182" width="41" height="10" font="6">13q13.3</text>
<text top="172" left="182" width="29" height="10" font="6">13qter</text>
<text top="159" left="236" width="40" height="10" font="6">13q21.1</text>
<text top="172" left="236" width="34" height="10" font="6">13q31.1</text>
<text top="159" left="292" width="46" height="10" font="6">13q21.32-</text>
<text top="172" left="292" width="29" height="10" font="6">13qter</text>
```

Figure 3.4: Corresponding XML output of a table in a PDF formatted paper

isolated from the table content. Footnotes can be useful as they sometimes provide extra information to help decipher cell values.

3.2.4 Transforming the raw table data into T^{phys}

The raw, unprocessed table data is not in a form that matches the criteria for T^{phys}, for instance, some cells can be empty. The reasons for having empty cells must be determined and the raw table must then be transformed into a T^{phys} compliant form.

- Some rows contain headings and the following rows contain sub-row headings. This can be determined from the XML left value of the text control information. The row heading would have a left value smaller than the left value of the content under the sub-row heading. Occasionally, some row contents are broken into multiple lines. To interpret this information correctly I need to reassemble them as a single line. My heuristic for this situation is: If the corresponding value

of the row in the first column (the row header cell) is empty then I join all values in the row with the previous row value, column by column. For example in Figure 3.7, the cell in row 3 column 1 is empty so this heuristic has been applied. "13qter" is joined with "13q13.3-" to make "13q13.3-13qter" which is the correct interpretation of these column headers.

- Under a category, the listed values can belong to two or three categories. For instance, under "Clinical Findings/Clinical Features" the contents can be "Phenotype" or "Disease Name" or it might fall into the "Abnormality" category.

  To categorize it correctly and enter it into the ontology under the appropriate concept, the category is verified with the help of other open source software to determine whether the cell refers to a phenotype name or a disease name or a syndrome name. The phenotype detection system used to determine the appropriate category is currently under development by M. Khordad member of the Cognitive Engineering Laboratory at UWO [26] . This phenotype detector combines the functionality of UMLS Metamap and OBO-Edit. The block diagram is given in Figure 3.5. Since none of the software used is able to provide the necessary information all of the time, if the category of any term is still undecided then the category of the term is provided manually by a human informant.

- On occasion, column headings are not available (for example, see Figure 3.6). To populate the ontology, a column heading is required. To solve the problem I take the caption of the table as a source of the information and try to provide an appropriate heading for the column. For Figure 3.6 the caption of the table suggests "Clinical Findings" as a heading for the missing column heading. To confirm this choice the next step is to check the values of that column and decide which values fall under this concept. In this table I encounter a further complication that I solve: "Clinical Findings" is a super concept of

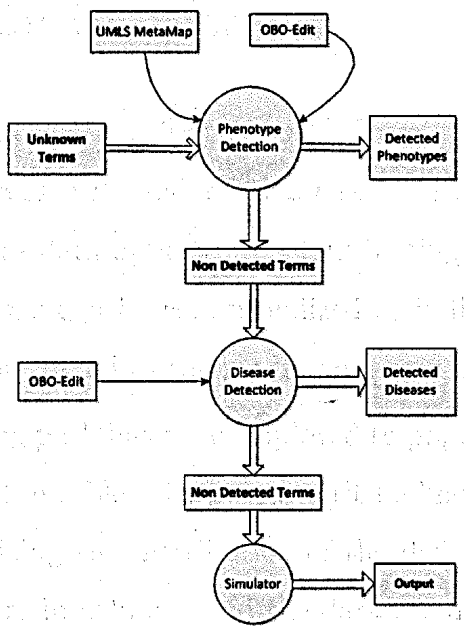

Figure 3.5: Block Diagram for Detection of unknown terms

Table 2: Clinical Findings in our Patient and Those in CFC and CS

	Our Patient	CFC %	Costello%[Gripp et.al., 2006a]
Polyhydramnios	-	46	87
FTT	+	100	100
Ulnar deviation	-	38	75
Hypotonia	+	69	72
CNS Abnormality	No data	31	27

Figure 3.6: Example of a table with an empty column heading. This table is taken from the paper Al-Rahawan et. al [1]

phenotype, syndrome or disease and Attribute of the subject (here the subject is "Patient" and the attributes are "age", "sex", "postnatal overgrowth", and "Birth Weight").

After isolating the table I store the table information in a form to preserve the important properties of the table as well as to process information easily. This information extraction phase is described next.

3.3 Information Extraction

To interpret the table information, first I use M. Hurst's concept of a reading path [16] to associate the *access cells* and the *data cells.* Data cells are characterized as terminating cells where data appears in a terminating role. Access cells are the remaining cells. These cells can be conceptualized as indices of the data cells. The reading of the table is considered as the reading of every data cell in the table.

This concept of reading path has been employed to populate the ontology. To construct the reading path, the table orientation must be known. The table orientation is determined by recognizing the distribution of the data. For example, for Figure 3.8, the main headings are in columns and the data are distributed for the subjects down the columns. However, for Figure 3.7, the main headings are in the rows and the data are distributed across the rows. I refer to the previous column oriented table as a vertical table and the latter row oriented table as a horizontal table.

3.3.1 Row-oriented (horizontal) and column-oriented (vertical) tables

Depending on which aspect of the table (row or column) is to be considered the main axis of the information, tables can be given an orientation. Based on the orientation of the data in the table, tables can be further categorized: Row-oriented tables and column-oriented tables. I will refer to the first type of table as a *horizontal table*, and the latter type as a *vertical table.* The table orientation affects the order that access cells are examined.

It is not a simple task to determine what is to be the table orientation, so for the purposes of this thesis, I have manually interpreted the orientation of the table based on the content of the caption. The two examples given below are intended to clarify this interpretation.

Table 3: Clinical features and size of deletion of the 12 patients with 13q monosomy.

Patients	1	2	3	4	5	6	7	8	9
Deleted segment	13q13.3-13qter	13q21.1-13q31.1	13q21.32-13qter	13q31.1-13q33.3	13q31.1-13qter	13q31.1-13qter	13q31.3-13q31.1	13q31.3-13q34	13q32.1-13qter
Size of deletion	70 mb	30 mb	47 mb	28 mb	34 mb	30 mb	10 mb	20 mb	18 mb
Sex	f	m	m	f	f	m	m	m	m
Child(c)/foetus(f)	f(33wg)	c(13m)	f(25wg)	f(24wg)	f(25wg)	f(26wg)	f(32wg)	f(23wg)	f(21wg)
IUGR	+	NK	+	-	+	+	-	-	+
Growth retardation	NK	+	NK	NK	NK	NK	NK	NK	NK
Microcephaly	+	-	-	-	+	+	-	+	+
Mental retardation	NK	+	NK	NK	NK	NK	NK	NK	NK
Brain anomalies									
Corpus callosum agenesis	-	-	+	-	+	NK	NK		
Holoprosencephaly	-	-	-	-	+	+	+	+	
Cerebellar vermis hypoplasia	+	-	+	-	-	NK	NK		
Digestive anomalies									
Common mesentary	-	-	-	+	-	-	-	-	-
Pancreas anomalies	-	-	-	+	-	-	-	+	-
Gall-bladder agencies	-	-	-	-	-	+	-	Hypoplasia	-
Abnormal lungs	+	NK	+	+	+	+	-	-	+

NK: not known, m: months, y: years, wg: weeks of gestation, f: female, m: male; arsa: aberrant right subclavian artery; dscv: double superior vena cava. (1) nasal bone hypoplasia.

Figure 3.7: Example of a Table from a set of tables based on that I engineered the proposed ontology. This table is reproduced from [24] with some rows and columns removed to fit the page. The rows are rearranged for addressing different types of problems.

3.3.2 Example of a Row-oriented (Horizontal) table

In a table in which the phenotype headings are the access cells for the rows and the data for the phenotypes are distributed across the rows, then this type of table is considered to be row-based or horizontal. The table presented in Figure 3.7 is a horizontal table where headings like Patient, Deleted Segment, size of deletion, and the various clinical features are the access cells for the rows and the corresponding value of the headings like patient ID value (1, 2, 3,...) and Deleted Segment value (13q13.3-13qter, 3q21.1-13q31.1,...) are distributed across the rows.

Table 4: HLXB9 Mutations Identified in the Study and Associated Phenotypes

Mutation Class	Mutation Position	Nucleotide Change	Amino Acid Change	Clinical Phenotype	Family or Patient No.
Missense	Homeobox	CrG, nt 4171	R247G	Hemisacrum, ARM, presacral mass, perianal abcess	3
Missense	Homeobox	TrG, nt 4900	W290G	Hemisacrum, ARM, presacral mass, rectovaginal fistula, neurogenic bladder	35
Missense	Homeobox	TrG, nt 4900	W290G	Hemisacrum, ARM, presacral mass, tethered cord	37

Figure 3.8: Example of a column-oriented table (This table is a shorter version of the original table)

3.3.3 Example of a Column-oriented (Vertical) table

If the main headings are in the column access cells and the data are distributed for the subjects down the columns then this type of table is referred to as column-oriented or vertical. For example, Figure 3.8 is an example of a vertical table. Headings like Mutation Class, Mutation Position, Nucleotide Change, etc. are the access cells for the columns and data for these columns are distributed down the columns.

3.3.4 Heuristics to determine table orientation

Now, I need some heuristics to distinguish horizontal and vertical tables automatically. After observing collection of tables I designed the following heuristics:

1. Common terms in the table caption and the access cells are an important source of information to distinguish horizontal tables and vertical tables. For instance, in Figure 3.8 the caption of the table is "HLXB9 Mutations Identified in the Study and Associated Phenotypes". In the table one column header contains "Mutation" in it. Since the column header contains the important term "Mutation" and the table caption contains the same term in one of its main noun phrases it is a vertical table. Similarly, in Figure 3.7, the caption of the table is "Clinical features and size of deletion of the 12 patients with 13q monosomy". It includes three important terms "patients", "deletions" and "clinical features" where all of them are present in the row header of the table. As the important

terms are found in the row header of the table, the table is a horizontal table.

2. Other heuristics can also be applied. Horizontal tables usually have many more columns than vertical tables; the first column heading starts with the Family or Patient terms followed by Age/Gender/Weight/Height attributes of the patient as row headings; and the remaining column headings contain attributes such as patient numbers like 1, 2, 3, ... or Identification numbers like CI, CII, ... rather than alphabetic attributes (see Figure 3.7). Very often, the data cell values are expressed as symbols like, "+", "-" or abbreviated words. For example, in Figure 3.7, NK stands for "not known".

3.3.4.1 Reading Paths for Oriented Tables

Hurst's concept of reading path has to correspond to the orientation of the table. Examples of reading paths are given below.

3.3.4.2 Maintaining Table Relations in the Ontology

Using the reading path concept I am able to reach a particular data cell. For example, considering Figure 3.8, the reading path for the data cell in row 2 column 2 (containing value Homeobox) is:

Mutation Class → Missense → Mutation Position → Homeobox

In this way, (I described it earlier) I am able to insert the value "Missense" and "Homeobox" and the relation between them in the ontology. However, the reading path concept is not able to retain the relation between "Homeobox" and the other cells in the row. To solve this problem I consider each row (containing data cells) instead of a particular cell and then combine the reading paths of all cells together taking common terms only once.

According to that concept, considering Figure 3.9, the reading path for the second row which contains the data cell:

Table 5: Dummy table

access cell c00 (access column heading 1)	access cell c01 (access column heading 2)	access cell c02 (access column heading 3)	....	access cell c0n (access column heading n)
access cell c10	data cell c11	data cell c12	....	data cell c1n
. . . .				
access cell cm0	data cell cm1	data cell cm2	....	data cell cmn

Figure 3.9: A dummy table indicating access cells and data cells (m=number of rows, n=number of columns)

> access column heading1 (cell c00) → access row value (cell c10) → access column heading 2 (cell c01) → data cell value(cell c11)→ access column heading3 (cell c02) → data cell value (cell c12) →... → access column heading-n (cell c0n) → data cell value (cell c1n).

If I apply this concept to the Table in Figure 3.8 I get:

> mutation class → missense → mutation position → homeobox → nucleotide change → crg, nt 4171 → amino acid change → r247g → clinical phenotype → hemisacrum, arm, presacral mass, perianal abcess → family or patient no. → 3.

Now I search for the first term "Mutation Class" in the ontology and once I find it I insert the second value "Missense" under the class "Mutation Class". Similarly, I enter other values from the reading path. In this way, I can populate the proposed ontology appropriately. At the same time, this reading path has confirmed the correctness of the design of the ontology. Moreover, from the reading path I know how data cells are connected with each other. I reflect this connection in the ontology by creating a new relationship and connecting the data cells with it. As an example, to preserve the relationship between data cells under "Mutation Class" and "Mutation Position" I build a relationship named "mClassmPos" and connect data cells with it. I illustrate this concept in Figure 3.10. Similarly I create all possible relationships (shown in Figure 3.10) to retain the relationship that exists among the fields in a table.

In the example, mClassmPos, mClassNucleo, MutClassAmino, PhenoMutPos, PhenoId, etc. all are relationship created to maintain the tabular relationships among the data cells in the row.

In the case of the horizontal table I need to change the order of the reading path.

> access column heading1 (cell c00) → access column heading2 (cell c01) → access row value (cell c10) → data cell value (cell c11) → access column heading3 (cell c02) → data cell value (cell c12) →... → access column heading n (cell c0n)→ data cell value (cell c1n)

Considering Figure 3.8, the reading path for row 2 is:

> Patients → 1 → Size of deletion → 70 Mb → Size of deletion → 30 Mb → Size of deletion → 47 Mb → Size of deletion → 28 Mb → Size of deletion → 34 Mb → Size of deletion → 30 Mb → Size of deletion → 10 Mb → Size of deletion → 20 Mb → Size of deletion → 18 Mb

Now I can populate the ontology as I did for table in Figure 3.8.

If there are n concepts in an ontology based on the design scheme I proposed then the number of relationships needed to maintain the tabular relationship in the table among the data cells in the row will be: $R_n = n-1+n-2+...+1 = n*(n-1)/2$.

3.3.4.3 Matching Access Cell Names with Ontology Concept Names

I observe that sometimes the column (or row) heading has a value, like "Classes of Mutation" or "Position of Mutation", instead of having values that directly match the name of the concept in the proposed ontology. In these cases I check the possible word variations. This transformation is further confirmed by observing the often highly stylized form of the data associated with that heading, making certain that it corresponds to the ontology concept that I have chosen.

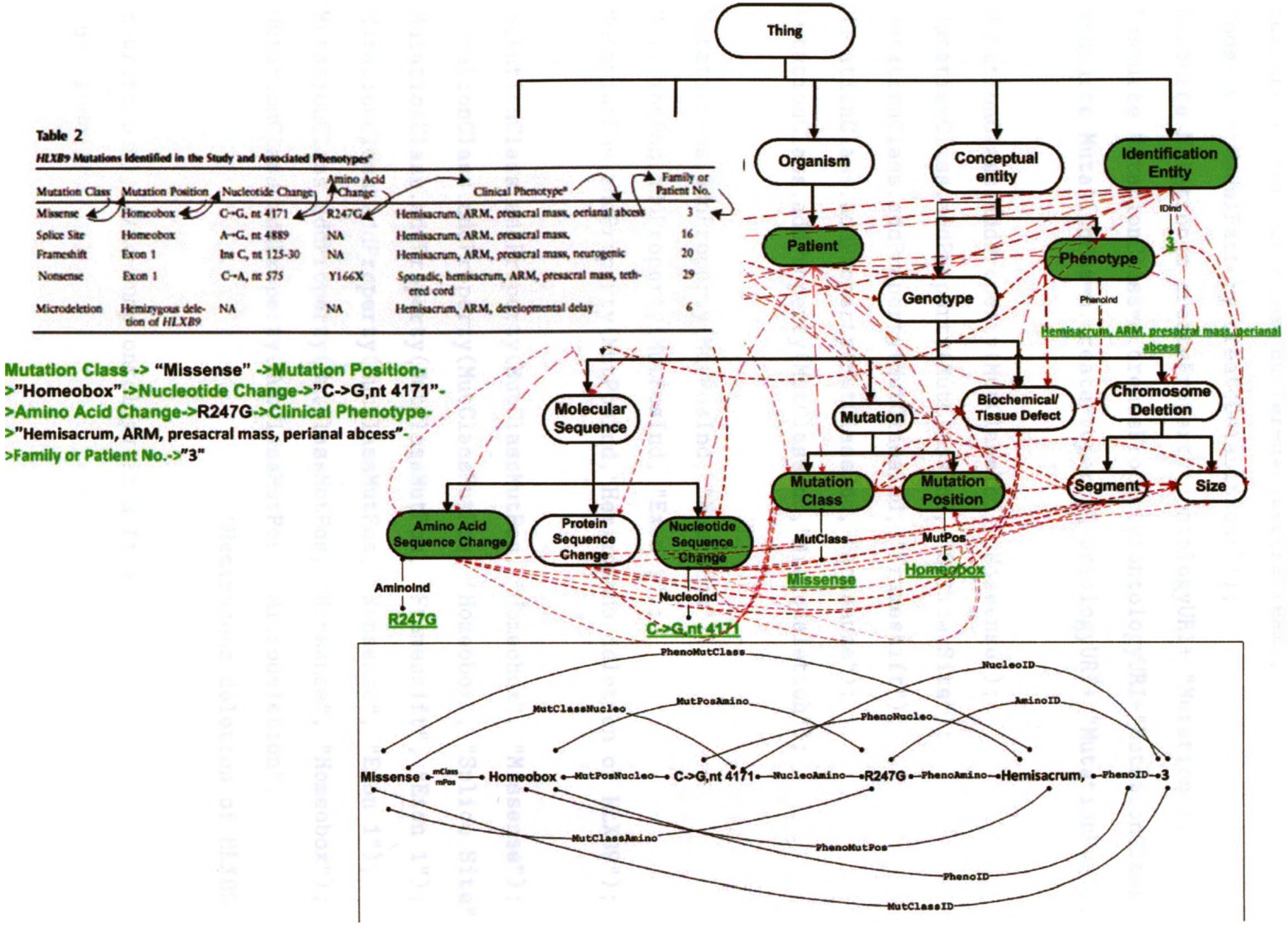

Figure 3.10: Population of the ontology

```
BufferedWriter out = new BufferedWriter(fstream);
Model m = ModelFactory.createDefaultModel();
Resource Mutation=m.createProperty( ontologyURI+ "Mutation");
Resource MutationClass=m.createProperty(ontologyURI+"MutationClass");
Resource MutationPos=m.createProperty( ontologyURI+ "MutationPos");

MutationClass.addProperty(MutClassInd, "Missense");
MutationClass.addProperty(MutClassInd, "Splice Site");
MutationClass.addProperty(MutClassInd, "Frameshift");
MutationClass.addProperty(MutClassInd, "Nonsense");
MutationClass.addProperty(MutClassInd,"Microdeletion");

MutationPos.addProperty(MutPosInd, "Homeobox");
MutationPos.addProperty(MutPosInd, "Exon 1");
MutationPos.addProperty(MutPosInd,"Hemizygous deletion of HLXB9");

MutationClass.addProperty(MutClassMutPos, "Homeobox", "Missense");
MutationClass.addProperty(MutClassMutPos, "Homeobox", "Splice Site");
MutationClass.addProperty(MutClassMutPos, "Frameshift", "Exon 1");
MutationClass.addProperty(MutClassMutPos, "Nonsense", "Exon 1");
MutationClass.addProperty(MutClassMutPos, "Nonsense", "Homeobox");
MutationClass.addProperty(MutClassMutPos, "Microdeletion",
                                    "Hemizygous deletion of HLXB9");

m.write(out); //to dump ontology in a file.
out.close();
```

Figure 3.11: Code for creating and dumping ontology in a file

3.4 Populating and Querying the Table Ontology

Furthermore to store information along with the relationship that exists between the cells in a particular table I employ Hurst's concept of a reading path [16] and change the sequence of reading path to store cell values along with the relationship from differently aligned tables. I described the whole process in the previous section. I created a XML file in Java to store the proposed ontology so that further retrieval and modification of the ontology is possible. To create and dump the ontology in a file format the code is shown in Figure 3.11. Here, I created a model m along with resources (to implement the concepts of the ontology) and properties (to implement all types of relationships existing in the table). These resources and the properties are added to the model m with the createProperty command. To save the whole model finally the model m is dumped into a file. Here the ontology is saved as an XML format in the file named *ontology_v1.xml*. If I open the file I can see the information saved in the format shown in Figure 3.12.

Finally, I developed a tool in Java that efficiently queries the ontology model and displays the output in a format which can be further used by other systems. The SPARQL query language for RDF graphs is used to query the ontology which is built upon the Resource Description Framework (RDF) using the Jena package.

For example, considering the example stated in Figure 3.11, to get information about all the values under Mutation Class the query in RDF is:

```
String queryString1 =
"PREFIX foaf: <http://modelOntology/tableOntology>"+
"SELECT ?MutClass"+
"WHERE {"+
 " ?x foaf:MutClass ?MutClass."+
 " }";
```

The result is listed in Figure 3.13. More query results will be shown in Chapter 4.

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:j.0="http://modelOntology/tableOntology#" >
 <rdf:Description rdf:about="http://modelOntology/tableOntology#MutationClass">
  <j.0:MutClassMutPos xml:lang="Exon 1">Frameshift</j.0:MutClassMutPos>
  <j.0:MutClassMutPos xml:lang="Hemizygous deletion of HLXB9">Microdeletion</j.0:MutClassMutPos>
  <j.0:MutClassInd>Splice Site</j.0:MutClassInd>
  <j.0:MutClassInd>Missense</j.0:MutClassInd>
  <j.0:MutClassInd>Frameshift</j.0:MutClassInd>
  <j.0:MutClassInd>Nonsense</j.0:MutClassInd>
  <j.0:MutClassInd>Microdeletion</j.0:MutClassInd>
  <j.0:MutClassMutPos xml:lang="Homeobox">Nonsense</j.0:MutClassMutPos>
  <j.0:MutClassMutPos xml:lang="Exon 1">Nonsense</j.0:MutClassMutPos>
  <j.0:MutClassMutPos xml:lang="Splice Site">Homeobox</j.0:MutClassMutPos>
  <j.0:MutClassMutPos xml:lang="Missense">Homeobox</j.0:MutClassMutPos>
 </rdf:Description>
 <rdf:Description rdf:about="http://modelOntology/tableOntology#MutationPos">
  <j.0:MutPosInd>Hemizygous deletion of HLXB9</j.0:MutPosInd>
  <j.0:MutPosInd>Exon 1</j.0:MutPosInd>
  <j.0:MutPosInd>Homeobox</j.0:MutPosInd>
 </rdf:Description>
</rdf:RDF>
```

Figure 3.12: Saved Ontology format in a file (this is a view of a ontology portion for smaller data)

MutClassInd
Missense
Splice Site
Frameshift
Nonsense
Microdeletion

Figure 3.13: Output of the query querystrng1

MutClassMutPos
"Homeobox"@Missense
"Homeobox"@Splice Site
"Frameshift"@Exon 1
"Nonsense"@Exon 1
"Nonsense"@Homeobox
"Microdeletion"@Hemizygous deletion of HLXB9

Figure 3.14: Output of the query queryString2

In some cases the situation might be a little more complicated. After doing a query from the result set, further string processing is needed to get the desired output. For example, from the ontology developed in Figure 3.11, if for a particular value of Mutation Class, say, Missense, I need to know the corresponding value of Mutation Position then it requires the following steps:

1. Find out the value the appropriate relation to call upon is MutClassMutPos.

2. Execute the query:

```
String queryString2 =
"PREFIX foaf: <http://modelOntology/tableOntology>" +
"SELECT ?MutClassMutPos" +
"WHERE {"+
"          ?x foaf:MutClassMutPos ?MutClassMutPos." +
"                              }";
```

3. Result set for the query (see Figure 3.14) for the above query.

4. Now to find out the record only related to the Missense, first I copy each line from the recordset into a string and then separate the string into two parts based upon the operator @. Say the first part is stored into variable A, and the

second part is stored into the variable B. Now if B ="Missense" then the value of A will be shown in the output.

3.5 Conclusion

This chapter describes the entire methodology of the thesis along with some examples. The next Chapter will describe the results of populating and querying the ontology.

Chapter 4

Results

This chapter reports on a further study of the table ontology populating software introduced in Chapter 3: the table isolation software, the table information transformation software, and the ontology population software. Of the 100 tables that were initially curated, those were used to design the software. I have collected eight new tables for testing purposes. The software has been tested on some of the 100 tables to show the correctness of the software and tested on the new tables to understand where the software fails to capture the full range of possible phenotype-genotype tables in the scholarly biomedical literature. The results of queries to the ontology are used to show the populated ontology.

Some of the material from Chapter 3 is purposefully repeated here to have all of the results in this chapter.

4.1 Creating T^{phys}

After isolating the table information the next step is to transform the table information into T^{phys} to facilitate interpreting the table information and to populate the

ontology properly. My approach to transforming the raw table information considers different factors like recognizing multiple and single line column headers of the table, distinguishing between sub-row heading and multiple line row content from the structurally different tables and handling it properly. In addition, problems like missing column headings, mainly in the first column, were encountered. In these cases the caption of the table is an important source of information for a suitable heading for the column. The details are given below.

- In Figure 4.1, the string "Cerebellar vermis hypoplasia" is broken across two rows and occurs under the sub-row heading "Brain anomalies". The approach to solve the problem is to distinguish between a sub-row heading and a multiple line row content. The part of the stored ontology is shown in Figure 4.2 where "Cerebellar vermis hypoplasia" is first joined together to form a single line and then attached with the sub-row heading "Brain anomalies". The approach of joining sub-row headings with the value can be useful for further processing. Similar things happen with the values of the row labeled "deleted segment" The value "13q13.3 -" and "13qter" should be joined together to form "13q13.3-13qter" to facilitate interpretation. In the stored ontology, shown in Figure 4.2, the values of "deleted segment" which spreads over two rows are joined together to be considered as a single phrase.

- In the table in Figure 4.3, column 1 does not contain a heading. To insert all the values of that column into the ontology, the column heading must be specified. The solution that has been found for this problem is to apply two heuristics: first an appropriate heading for column 1 can be taken from the caption of the table and second, the concept type of the data in that column can be considered as a clue to the appropriate column heading. Parsing the table caption produces "Clinical Findings" and "those" as the main noun phrases conjoined by the conjunction "and". The pronoun "those" refers to "Clinical

Findings". The prepositional phrases correspond to other column headings. So, "Clinical Findings" is an appropriate heading for column 1. This parsing of the table caption has not been automated. Automatic parsing has been left for future work. The string "Clinical Findings" is supplied manually to the software.

Additionally, if I take a look at the data under the missing column heading, then all the values except "CNS Abnormality" are phenotypes which are appropriate examples of "Clinical Findings". "CNS Abnormality" can be considered as a cause associated with a syndrome. So the appropriate heading for the column could be "Clinical Findings" or "Clinical Features". Since "Clinical Findings" is suggested by both heuristics, this is the heading chosen.

- In Figure 4.5, a fifth column cell content breaks across five rows where usually cell contents break across two/three rows. Query Y shows that my heuristic approach correctly forms a single line from multiple lines.

4.1.1 Populating the ontology

The table ontology is one of the main contributions of this work. I engineered a table ontology that incorporates some new terms (compared to the UMLS ontology) to include relationships that exist in these tables (e.g. subjects have an age, or certain important terms like genotype, phenotype are concepts not found in the UMLS ontology).

After coming up with a blueprint of the ontology, the main task is to populate the ontology properly. Populating the ontology involves certain tasks such as: matching the headings of the columns and the rows to the ontology concepts, inserting the appropriate data values in the reading path under those concepts, and retaining the

Table 3: Clinical features and size of deletion of the 12 patients with 13q monosomy.

Patients	1	2	3	4	5	6	7	8	9
Deleted segment	13q13.3-13qter	13q21.1-13q31.1	13q21.32-13qter	13q31.1-13q33.3	13q31.1-13qter	13q31.1-13qter	13q31.3-13q31.1	13q31.3-13q34	13q32.1-13qter
Size of deletion	70 mb	30 mb	47 mb	28 mb	34 mb	30 mb	10 mb	20 mb	18 mb
Sex	f	m	m	f	f	m	m	m	m
Child(c)/foetus(f)	f(33wg)	c(13m)	f(25wg)	f(24wg)	f(25wg)	f(26wg)	f(32wg)	f(23wg)	f(21wg)
IUGR	+	NK	+	-	+	+	-	-	+
Growth retardation	NK	+	NK	NK	NK	NK	NK	NK	NK
Microcephaly	+	-	-	-	+	+	-	+	+
Mental retardation	NK	+	NK	NK	NK	NK	NK	NK	NK
Brain anomalies									
Corpus callosum agenesis	-	-	+	-	+	NK	NK		
Holoprosencephaly	-	-	-	-	+	+	+	+	
Cerebellar vermis hypoplasia	+	-	+	-	-	NK	NK		
Digestive anomalies									
Common mesentary	-	-	-	+	-	-	-	-	-
Pancreas anomalies	-	-	-	+	-	-	-	+	-
Gall-bladder agencies	-	-	-	-	-	+	-	Hypoplasia	-
Abnormal lungs	+	NK	+	+	+	+	-	-	+

NK: not known, m: months, y: years, wg: weeks of gestation, f: female, m: male; arsa: aberrant right subclavian artery; dscv: double superior vena cava. (1) nasal bone hypoplasia.

Figure 4.1: Example of a Table from a set of tables based on that I engineered the proposed ontology. This table is reproduced from [24] with some rows and columns removed to fit the page. The rows are rearranged for addressing different types of problems.

relationships among the cells in the rows and columns of the table so that future querying of the ontology is as informative as the original table.

The ontology population procedure is different for Horizontal tables and Vertical tables. Some techniques to distinguish horizontal and vertical tables have been introduced in Chapter 3. After distinguishing the Horizontal tables from the Vertical tables, different ontology populating procedures are invoked to populate the ontology. Details of these procedures are provided in the following sections.

Figure 4.2: Ontology stored in XML format

Table 4: Clinical Findings in our Patient and Those in CFC and CS.

	Our Patient	CFC %	Costello%[Gripp et.al.,2006a]
Polyhydramnios	-	46	87
FTT	+	100	100
Ulnar deviation	-	38	75
Hypotonia	+	69	72
CNS Abnormality	No data	31	27

Figure 4.3: Example of a Table with missing Column heading

4.1.1.1 Populating the ontology with Horizontal table information

As described in Chapter 3, a Horizontal table is a row-oriented table where the main headings are in the rows and the data are distributed across the rows. Once the table isolation and curation of the table text is done, the next step is to find the appropriate

ontology concept under which to store the information. In Figure 4.1, to populate the ontology the first item to consider is "Patients".

In the proposed ontology there is a concept named Patient under the Organism concept which also contains other concept like Age, Gender, Weight as a child. However, in the case of Figure 4.1, here the patient concept more likely acts as an identification entity for the table data. So, the appropriate position in the ontology for this data is Identification Entity. For the other row headings like Sex, the ontology has a concept named Gender to accomodate this information. So after the extracting row header the next step is to perform several string matching operations particularly considering synonyms of the name used in the concept of the ontology. The partial stored ontology of the table is shown in Figure 4.2.

Similarly, the values of "Deleted segment" and "Size of Deletion" can be stored into the ontology. However, in the case of the other row heading IUGR, Growth retardation, Microcephaly, etc., it is first necessary to distinguish these terms and then select the appropriate ontology. A several procedure is used to determine unknown terms described in Chapter 3. After detecting a term whether it is phenotype or genotype or Disease or syndrome then it is easy to insert into the ontology.

It is necessary to store not only the data but the relationship among the cells to retain the relationship that exists in the tabular form. In this connection an additional property has been created to establish the connection among cells. In Figure 4.1 the stored ontology representation of the relationship between Patient with Size of Deletion and Patient with Deleted Segment is shown in Figure 4.7. The relationship between Patient and Size of Deletion is referred to "ChromDelSizeId". It connects all data of Patient with all the data of Deleted Segment maintaining the relationship that exists between them in the table (Red arc on the Figure 4.7). Similarly, the connection (Green arc) between Patient and the Size of Deletion is established by "ChromDelSegId". Moreover, to facilitate correctness of query to the ontology a

direct connection (Blue arc) between Size of Deletion and Deleted Segment is also created.

In the case of "Abnormal lungs", "Growth retardation", "Microcephaly", "Mental retardation", first it has to be determined in which classes (Phenotype, Disease, Syndrome, or Abnormality) these terms belongs to. The unknown term detection system described in Chapter 3 confirms Microcephaly, mental retardation, and growth retardation to be Phenotypes. However, Abnormal lungs is neither a Phenotype nor a Syndrome/Disease. It falls into a group named Abnormality in the Table Ontology.

Now from the structure of the table, it appears that to store the tabular relationship between Microcephaly and Deleted segment only relationship between these two terms needed to be created. However, in the implementation phase the scenario is different. As Microcephaly falls into the Phenotype group so the relationship between Phenotype and Deleted Segment should be used here. Furthermore, these Phenotypes and Deleted Segments have their own values like Microcephaly and 13q31.3-13q34 respectively and they have a value in common to indicate that whether this particular Phenotype has any relation with Deleted Segment or not. So to accomodate this extra information an additional relationship (property) has to be created. Furthermore, to retain the relationship among cells as they are in table, this additional information is also needed to be mapped with Phenotypes and Deleted Segment information. To implement this, creating a concept instead of a relationship is necessary. To represent these type of values I considered this as clinical findings and store these values to the ClinicalFinding concept. A detailed procedure of mapping among Phenotypes and Deleted segment and their corresponding clinical findings (that this Phenotype has what type of relation with the Deleted segment) are described below:

1. Microcephaly is added as a Phenotype by invoking the relationship PhenoInd. This relationship is added with concept Phenotype. The code is shown in below:

```
Phenotype.addProperty(PhenoInd,"Microcephaly");
```

2. Similarly, 13q13.3-13qter is added with Deleted Segment.

```
ChromDelSegment.addProperty(ChromDelSegInd,"13q13.3-13qter");
```

3. Now a relationship named PhenoChromDelSeg is used to create connection between Microcephaly and 13q13.3-13qter.

```
Phenotype.addProperty(PhenoChromDelSeg,"Microcephaly",
"13q13.3-13qter");
```

4. Now Microcephaly and 13q13.3-13qter map to an additional value which represents whether this particular Phenotype has what type of relation with Deleted Segment. Here in this case from the value of Figure 4.1, Microcephaly has + relation with Deleted Segment. So to implement the whole idea two additional relationships are required to map between three concepts. ClFindPheno relationship is created to map between values of Clinical Finding and Phenotype. This is added with the concept ClFinding.

```
ClFinding.addProperty(ClFindPheno,"+","Microcephaly");
```

Similarly, to retain the relationship between Clinical Finding and Deleted Segment ClFindChromDelSeg is used and added with the concept ChromDelSegment.

```
ChromDelSegment.addProperty(ClFindChromDelSeg,"+",
"13q13.3-13qter");
```

However, this approach has one problem. In Figure 4.1 and the above example, the listed values are the partial values stored under that relationship listed in the header in Figure 4.5. Now if I give a closer look then it is clear that Ocular anomalies Retinal dysplasia is mapped with +, - and NK (see Figure 4.1). Moreover, at the same

time it is related to the different deleted segments say13q13.3-13qter, 13q21.1-13q31.1, 13q21.32-13qter. Similarly, "Deleted segment" has different clinical value mapping with them. So if anyone tries to come up with the relationship among a particular phenotype, deleted Segment and Clinical Findings then it is difficult to figure out from the individual pair of relationship. For an example if the relationship among Phenotype (say, Microcephaly), Deleted Segment 13q13.3-13qter, and the clinical Finding is needed to be retrieved then from the information listed in Figure 4.4, for the Phenotype Microcephaly and for the Deleted segment 13q13.3-13qter the possible values listed here is +, -,NK where from the Figure 4.1 that should be +. The problem arises because the pair of relationship ClinicalFinding and Deleted Segment does not involve the third term (here phenotype) and one Deleted Segment may contain all possible values of Clinical Findings.

To solve this problem I apply a simple technique of adding an automatically generated ID with every entry. That is, when for a particular phenotype, its corresponding deleted segment and Clinical Findings get updated in the ontology, I associate a ID with the phenotype, Deleted segment and clinical findings at the same time. So when this information gets updated as a form of relationship pair, this associated ID works as an Identifier which helps us later to find out the phenotype, deleted segment and phenotype which got updated at the same time.

For example, taking the same example as above, the steps with the modified codes are shown below:

1. To add any Phenotype by invoking the relationship PhenoInd. This relationship is added with concept Phenotype.

   ```
   Phenotype.addProperty(PhenoInd,"Microcephaly");
   ```

2. To add any deleted segment the value is added with Deleted Segment.

```
ChromDelSegment.addProperty(ChromDelSegInd,"13q13.3-13qter");
```

3. Now a relationship named PhenoChromDelSeg is used to create the connection between Phenotype and Deleted Segment with the Phenotype ID attached with it. Here, *pheno_uni_index* is an automated generated number used as a Phenotype ID started from zero and after each entry it increases by 1. Here P stands for Phenotype.

```
Phenotype.addProperty(PhenoChromDelSeg,
"Microcephaly"+"*P"+pheno_uni_index,"13q13.3-13qter"
+"*P"+pheno_uni_index);
```

4. Now Phenotype and Deleted Segment map to an additional value which represents whether this particular Phenotype has what type of relation with Deleted Segment. So to implement the whole idea two additional relationships are required to map between three concepts. ClFindPheno relationship is created to map between values of Clinical Finding and Phenotype. This is added with the concept ClFinding.

```
ClFinding.addProperty(ClFindPheno,rrpath.get("+" + "*P"+
pheno_uni_index,"Microcephaly"+"*P"+pheno_uni_index);
```

Similarly, to retain the relationship between Clinical Finding and Deleted Segment ClFindChromDelSeg is used and added with the concept ChromDelSegment.

```
ChromDelSegment.addProperty(ClFindChromDelSeg, "+"  +"*P"+
pheno_uni_index, "13q13.3-13qter"+"*P"+pheno_uni_index);
```

then identifies is increased each time by one.

```
pheno_uni_index++;
```

Deleted Segment, Phenotype	Phenotype,ClinicalFinding	ClinicalFinding,Deleted Segment
13q13.3 -13qter,Microcephaly	Microcephaly,+	NK,13q13.3 -13qter
13q13.3 -13qter,Brain anomalies Corpus callosum agenesis	Brain anomalies Corpus callosum agenesis,+	+, 13q13.3 -13qter
13q13.3 -13qter,Ocular anomalies Retinal dysplasia.	Ocular anomalies Retinal dysplasia,+	-,13q13.3 -13qter
13q21.1 -13q31.1,Ocular anomalies Retinal dysplasia.	Ocular anomalies Retinal dysplasia,-	NK,13q21.1 -13q31.1
13q21.32 -13qter,Ocular anomalies Retinal dysplasia.	Ocular anomalies Retinal dysplasia, NK	+,13q21.32 -13qter

Figure 4.4: Data holds by the relationships

Table 5: Clinical features of the well-established or atypical cases of steinfeld syndrome.

Clinical signs	1	2	3	4	5b
sex	Female	Male	Female	Male	Female
child(c)/foetusC	F	F	F	F	F
Familila History	+ (AD?)	+ (AD?)	-	-	-
Cerebral malformation	Holoprosencephaly, pachygyria	Alobar holoprosencephaly	Semilobar holoprosencephaly	Holoprosencephaly/ frontal encephalocele/ lissencephaly/ brainstem and cerebellar hypoplasia	Semilobar holoprosencephaly
Ocular anomalies	?	Cyclopia	-	Anophthalmia	Right microphthalmia

ToF:Tetralogy of Fallot; DSCV: Double Superior Vena Cava.

Figure 4.5: Example of a table where row contents are broken down into multiple lines. This table is reproduced from [24] with some rows and columns removed to fit the page

In case of disease instead of phenotype the whole process is the same, just that there is another Identifier *disease_uni_index* is used with same nature that of *pheno_uni_index*. The sequence of characters attached with this identifier is "*D" instead of "*P".

Similarly, for Abnormality the identifier *abnormal_uni_index* is used and the sequence of character used with *abnormal_uni_index* is "*A".

4.1.1.2 Populating the ontology with Vertical table information

I will now describe the population of the ontology for vertical tables. In Figure 4.6 for an example, focussing on the headings"Mutation Class", "Amino Acid Change terms", the corresponding populating ontology steps is given below:

1. To add any Mutation Class by invoking the relationship MutClassInd. This relationship is added with concept Mutation Class.

   ```
   MutationClass.addProperty(MutClassInd,"Missense");
   ```

2. To add any Amino Acid Change by invoking the relationship AminoAcidInd. This relationship is added with concept AmiAcidChange (Amino Acid Change).

   ```
   AmiAcidChange.addProperty(AminoAcidInd,"R247G");
   ```

3. Now a relationship named MutClassAmino is used to create a connection between Amino Acid Change and Mutation Class. This relationship can be added with either MutationClass or Amino Acid Change.

   ```
   MutationClass.addProperty(MutClassAmino,"Missense","R247G");
   ```

The procedure is the same for all other column headings for vertical tables.

In Figure 4.6, the number of relationships used to map between two cells is 21, creating relationships between one cell with every other cell. If there are n columns then the number of relationships that need to be created as: $n-1+n-2+...+1 = n*(n-1)/2$.

Table 6: HLXB9 Mutations Identified in the Study and Associated Phenotypes.

Mutation Class	Mutation Position	Nucleotide Change	Amino Acid Change	Clinical Phenotype	Family or Patient No.
Missense	Homeobox	CrG, nt 4171	R247G	Hemisacrum, ARM, presacral mass, perianal abcess	3
Missense	Homeobox	TrG, nt 4900	W290G	Hemisacrum, ARM, presacral mass, rectovaginal fistula, neurogenic bladder	35
Missense	Homeobox	TrG, nt 4900	W290G	Hemisacrum, ARM, presacral mass, tethered cord	37

Figure 4.6: Example of a vertical Table taken from the set of tables on which I engineered the proposed ontology

Figure 4.7: Ontology stored in XML format

4.1.1.3 Some queries to the ontology and their results

Query X: List the phenotypes and their corresponding Clinical Finding for:

A. For all Deleted Segment.

B. Size of deletion greater than 20 Mb.

Query Y: Considering Figure 4.5

Show records of all patients where Clinical features is narrow down to Cerebral malformation.

Discussion of the results:

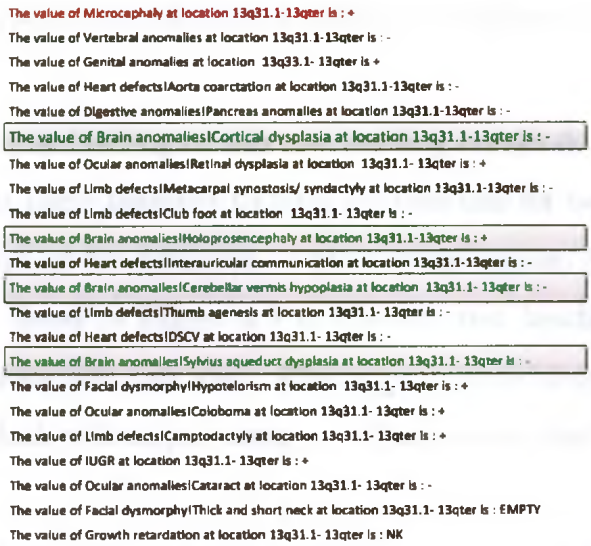

Figure 4.8: Result of query X A

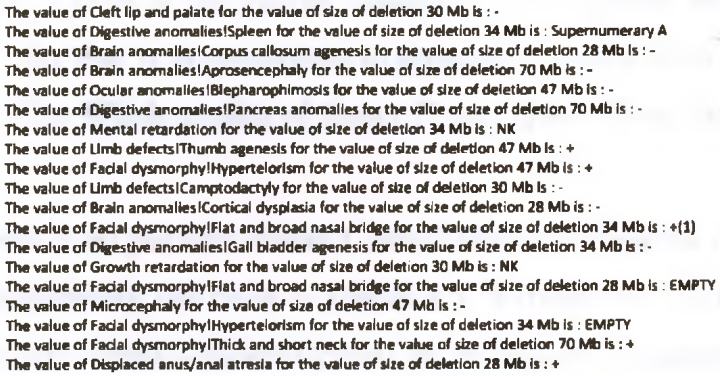

Figure 4.9: Result of query X B

Query X.

In Figure 4.1 the partial result of the query is listed in Figure 4.8 and Figure 4.9, the rest are in the Appendix A. A: In Figure 4.1 the entry for Microcephaly at location 13q31.1 to 13qter is +. Moreover if a category contains a sub category, then it should always be recognized and attached with the super category so that table information interpretation can be done more accurately. In the example above, the category Brain anomalies is attached with the sub category Holoprosencephaly,

Cerebellar vermis hypoplasia and Sylvius aqueduct dysplasia as shown in the Figure 4.8.

Moreover, if any row heading or sub row heading breaks down into multiple lines it is necessary to put them together to form a single line for better interpretation of table information and/or for detection of phenotype/disease/Abnormality properly. In the query result listed in Figure 4.8 in the sub row heading Cerebellar vermis hypoplasia breaks down into two rows. This term is joined together to form a single line and then attached with super category "Brain anomalies" (Highlighted in the figure).

B: In Figure 4.9 the query result listed some values of Phenotypes and their corresponding clinical findings where size of Deletion is greater than 20 Mb.

In the results of both of the queries categories are combined with sub categories with symbol "!", so that it is convenient to separate category from the sub category later if needed. The Whole results of Query X and Query Y are listed in Appendix A.

Query Y: Records of all patients where Clinical features is narrow down to Cerebral malformation considering the table in Figure 4.5. Patient No. : 4 Clinical Findings : Holoprosencephaly/ frontal encephalocele/ lissencephaly/ brainstem and cerebellar hypoplasia

Patient No. : 3 Clinical Findings : Semilobar holoprosencephaly

Patient No. : 2 Clinical Findings : Alobar holoprosencephaly

Patient No. : 7 (Patient 6) Clinical Findings : Alobar holoprosencephaly

Patient No. : 1 Clinical Findings : Holoprosencephaly, pachygyria

Patient No. : 6 Clinical Findings : Atelencephaly

Patient No. : 5 Clinical Findings : Semilobar holoprosencephaly

Here the clinical record of patient 4 which breaks down into seven rows joined together and formed single line.

Testing Table 1: Summary of 77 different COL2A1 mutations identified in a series of 100 affected individuals.

Patient ID	Age (years)	Score	Exon/ intron	cDNA	Protein	Mutation type	Mutation effect
1	46	15		del COL2A1	del COL2A1	Large deletion	Deletion[19]
2	54	13	1	c.211233dup	p.Glu79ThrfsX2	Duplication	Frameshift
3	58	17	02	c.264276del	p.Cys89SerfsX24	Deletion	Frameshift
4	4	6	IVS 04	c.342+1G4A	p.Asp114Ile115insIleSerAlaAsn TyrSerHisProValLeuGlnLeuLeuX14	RNA processing	Insertion with premature stop codon
5	42	17	IVS 06	c.430-1G4C	p.Gly144ValfsX54p.Gln125Gly126 insArgGluGlyGluAsnLeuPheLeuArgPro a PheLeuAlaAlaGlnValThrAspLeuX20 p.Lys143Asn178delExon7a	RNA processing a	Frameshift, insertion with premature stop codon, exon deletion
6	6	6	07	c.492delT	p.Gly165ValfsX34	Deletion	Frameshift
7	3	6	09	c.625C4T	p.Arg209X	Nonsense	Premature stop codon
8	6	11	09	c.625C4T	p.Arg209X	Nonsense	Premature stop codon

Figure 4.10: Example of a testing Table taken from the set of new examples.

4.1.2 Testing Tables

The proposed ontology is further verified by populating a new set of tables. Here I will describe some examples. What follows is a discussion of the problems encountered when the table tool performance was investigated using these previously unseen tables.

4.1.2.1 Table 1

The table in Figure 4.10 is reproduced from Hoornaet et. al [11] with some rows removed to fit the page. While populating the ontology the following problem was encountered:

- In Figure 4.10 the "cDNA" and "score" column headers are new to the system. Further analysis of the term "cDNA" indicated that the values in the cDNA column represent the changes in the DNA that cause mutation. So, the decision was to enter this value under the MUTATION concept of the ontology. Moreover, score value is accomodated under EXPERIMENTAL FINDING concept.

The repaired ontology is populated with this table information and queries are made to verify the correctness of the population. Queries for each column of the table

Testing Table 2: Mutations in the MECP2 gene found in Danish patients with Rett syndrome

Nucleotide changes	Type of mutation	No. of patients	Phenotypeb
76delC	Frameshift	1	Cl
316 C ! T	R106W	1	Cl
397 C ! T	R133C	2	At, PSV
423 C ! G	Y141X	2	2 Cl
473 C ! T	T158M	7	5 Cl, 2 At
502 C ! T	R168X	4	4 Cl
763 C ! T	R255X	5	5 Cl
766dup14	Frameshift	1	Cl
808 C ! T	R270X	3	2 Cl, 1 nk
863-881del	Frameshift	1	nk
880 C ! T	R294X	6	5 Cl, 1 At
905 C ! T	P302L	1	Cl
916 C ! T	R306C	3	3 Cl
1144-1193del	Frameshift	1	Cl
1152-1189del	Frameshift	1	Cl
1156-1199del	Frameshift	1	Cl
1158-1201del	Frameshift	1	At
1163-1188del	Frameshift	1	At
1164-1217del	Frameshift	1	nk
1169-1197del	Frameshift	1	Cl

Figure 4.11: Example of a testing Table taken from the set of new examples.

stored in the ontology are shown in Appendix C. The query results were checked against the paper [11] and verified manually.

4.1.2.2 Table 2

The table that is in Figure 4.11 is reproduced from Nielson et. al [21]. No problems were encountered while populating the ontology. Results of queries to retrieve data from every column are listed in Appendix D and verified manually.

4.1.2.3 Table 3

The table that is in Figure 4.12 is reproduced from Bonuccelli et. al [3]. While populating the ontology some questions arise about the correct interpretation of column headings. In Figure 4.12 the column headers contain terminology which is misleading in one case and requires some analysis in some others. These problems are discussed below:

- In the first column heading, "Phenotypes" indicates that the information listed below should be different types of phenotypes. However, the data content refers to the intensity of phenotypes rather than the name of phenotypes. After having a close look at the document, in the introduction section of the paper there is a line which contains relevant information about the problem. "A broad spectrum of clinical phenotypes has been observed ranging from the mild form with late onset and absent or moderate mental retardation to the severe neuronopathic form with early death". From this information, the related phenotype with the disease can be "mental retardation". Using online sources, it was further verified that many phenotypes have intensity in the phenotype name. However, in the online source, many phenotypes are discussed, none of which are the topic of the paper. It is not a straightforward exercise to choose the correct phenotype.

- Consider the second column heading "Genotype" in Figure 4.12, the values listed below are mutations which can be further confirmed by reading the paper. In the proposed table ontology MUTATIONS is hierarchically under GENOTYPES so, from this point of view, considering the values in the column under the heading label "genotype" as mutations makes sense.

- In the column labeled "ER site +/-" the data cell values are locations on a DNA molecule containing specific sequences of nucleotides, which are recognized by *restriction enzymes*. So the information in this column can be inserted under the NUCLEOTIDES SEQUENCE concept in the ontology.

All other column values can be inserted into the ontology without requiring additional analysis. How to automatically search for and come up with an appropriate label for an unknown term is beyond the scope of the thesis; however, the knowledge sources do exist.

Testing Table 3: Patients' data and details about the corresponding mutations

Phenotype	Genotype	Mutation description	Exon	Nucleotide change	ERa site +/-	Protein alteration
Severe	K347T	missense	VIII	$AAA > ACA$	+MaeIII	$Lys > Thr$
Intermediate	N265I	missense	VI	$AAC > ATC$	+FokI	$Asn > Ile$
Severe	473delTCC	in-frame deletion	III	$TCC > -$	+HphI	Loss of a serine at codon 117
Severe	533delTT	not in-frame deletion	III	$TTT > -T$	+Alw26I	Chain termination

Figure 4.12: Example of a testing Table taken from the set of new examples.

Testing Table 4: Oligonucleotides used for PCR directed site mutagenesis.

Mutations	Oligonucleotide sequences
K347T	5P:GGAGAATGGGCCACATACAGCAATTTTG:3P
473delTCC	5P:GCTGGAAACTTC__ACCATCCCCCAG:3P
533delTT	5P:GTGGGAAAAGTC_TCACCCTGGGAT:3P
N265I	5P:GTGGCCTACATCCCCTGGATGG:3P
Sel:XhoI	5P:GTGCCACCTGGCTCGAGATTGATTATTGA:3P

Figure 4.13: Example of a testing Table taken from the set of new examples.

4.1.2.4 Table 4

The table that is in Figure 4.13 is reproduced from Bonuccelli et. al [3]. The information presented here can be populated into the ontology without any confusion in the concepts named MUTATION and PRIMER SEQUENCE.

4.1.2.5 Table 5

The table that is in Figure 4.14 is reproduced from Bonuccelli et. al [3]. In this table under the heading "Constructs", the column contains mutation and non-mutation values. The term "Constructs" refers to the product of transfection, where transfection is the process of deliberately introducing nucleic acids into cells. What is at issue here is that data cell values do not determine a unique concept in the ontology. It is obvious that uniqueness of concept is not guaranteed for data cell values such as integers, but it came as a surprise for values such as those found in Figure 4.14. One anticipated heuristic for determining the ontology concept of an unknown column header is to use the common concept of the column data values, but this heuristic needs to be evaluated.

Testing Table 5: IDS activity in COS 7 transfected with wild-type and mutant cDNAs.

Constructs	IDS activity (U/mg protein)
K347T	12.7
473delTCC	12.2
533delTT	13
N265I	29.6
Wild-type	394.8
Untransfected cells	12.3

Figure 4.14: Example of a testing Table taken from the set of new examples.

In order to populate the ontology with values of the column labeled "Constructs", I have added a new concept named CONSTRUCT under the concept conceptual entity. However, IDS activity can be inserted in the EXPERIMENTAL FINDING concept. A relationship can be created between CONSTRUCT and EXPERIMENTAL FINDING to retain the relationship between the values of construct and IDS activity in the table.

4.1.2.6 Table 6

Figure 4.15 is reproduced from Psoni et. al [23]. The proposed ontology is populated with all of the terms listed in the table. Two queries are executed to retrieve information and verified it manually to ensure proper population.

Unanticipated errors from the *TableSeer* table isolation software still occur. When encountering a " ' " in the table cell data, *TableSeer* produces an error code ('H11032') that is placed in its output. (In Figure 4.15). After doing query the query, sequence is converted to the original character " ' ". Moreover, the column header "Tm" actually refers to "temperature". So, this is further evidence that for the correct population of ontology more analysis of the column headers (" (°C)" is a good indicator of temperature) or a good list of synonyms is required.

Query VI A: List all primer sequences and all fragment values of the Exon 4.

Result: This is the complete result of the query VI A.

The primer sequence is 5'-CAG TTC CTG GGA AGC TCC TTG TCA AGA T-3' and fragment value is 4.2F for exon value of 4.

Testing Table 6: The primers and PCR conditions designed for the analysis of exons 3 and 4 of the MECP2 gene

Exons Fragment Primers Product size Tm (°C).

Exons	Fragment	Primers	Product size	Tm (°C)
3	3.1F	5H11032:AAG ATC TGA GTG TAT GAT GGC CTG GG:3H11032	428bp	60
3	3.1R	5H11032:TTT GCT TAA GCT TCC GTG TCC AGC:3H11032		60
3	3.2F	5H11032:AAG AGA AAG AGG GCA AGC ATG AGC :3H11032	405bp	60
3	3.2R	5H11032:AAG CAC ACC TGG TCT CAG TGT TCA :3H11032		60
4	4.1F	5H11032:CAG TTT GTC AGA GCG TTG TCA CC A CCA T:3H11032	626bp	62
4	4.1R	5H11032:TGA CGG AGT ACG GTC TCC TGC ACA GAT:3H11032		62
4	4.2F	5H11032:CAG TTC CTG GGA AGC TCC TTG TCA AGA T:3H11032	616bp	62
4	4.2R	5H11032:TGA CTC CTC TGG GCA TCT TCC CTC TTT:3H11032		62
4	4.3F	5H11032:CAG TGG GAA AGG ACT GAA GAC CTG TAA G:3H11032	566bp	60
4	4.3R	5H11032:TGA CCA GTT AAT CGG GAA GCT TTG TCA G:3H11032		60

Figure 4.15: Example of a testing Table taken from the set of new examples.

The primer sequence is 5'-TGA CTC CTC TGG GCA TCT TCC CTC TTT-3' and fragment value is 4.2R for exon value of 4.
The primer sequence is 5'-TGA CCA GTT AAT CGG GAA GCT TTG TCA G-3' and fragment value is 4.3R for exon value of 4.
The primer sequence is 5'-TGA CGG AGT ACG GTC TCC TGC ACA GAT-3' and fragment value is 4.1R for exon value of 4.
The primer sequence is 5'-CAG TGG GAA AGG ACT GAA GAC CTG TAA G-3' and fragment value is 4.3F for exon value of 4.

Query VI B: Show the Temperature and Product size of the fragment containing value 3.2F.

The result of the query is: The temperature and product size is 60 and 405bp respectively for fragment value 3.2F. This refers to the value of 3rd row in the table.

4.1.2.7 Table 7

Figure 4.16 is reproduced from Concolino et. al [6]. In this table all columns except "Activity in vitro 17:OHP/progesterone" can successfully be mapped into the

proposed ontology. However, according to wikipedia 17:OHP/progesterone or 17-Hydroxyprogesterone (17-OH progesterone or 17OHP) is a C-21 steroid hormone which is produced at the time of the synthesis of glucocorticoids and sex steroids. In the proposed ontology a concept named STEROID is under the subtree CHEMICAL VIEWED STRUCTURALLY → LIPID. So all values under the column 17:OHP/progesterone are stored under the concept steroid.

Query VII A: Show all the clinical phenotypes and the corresponding mutation values taken from the reference paper Tardy et al.

Result : For the clinical phenotype SW the corresponding mutation value is p.L167P
For the clinical phenotype SW the corresponding mutation value is p.G292D
For the clinical phenotype SV the corresponding mutation value is p.E320K
For the clinical phenotype CL the corresponding mutation value is p.Y59N
For the clinical phenotype CL the corresponding mutation value is p.M1V
For the clinical phenotype NC/SV the corresponding mutation value is p.R233K
For the clinical phenotype NC/SV the corresponding mutation value is p.R369W
For the clinical phenotype NC the corresponding mutation value is p.I230T

Query VII B: Show the gene and mutation values for all clinical phenotypes.

Result: This is the partial result of the query. The full result is listed in Appendix B.1.

For the clinical phenotype NC/SV the corresponding mutation value and gene expression is p.R233K and $g.1586G > A$ respectively.
For the clinical phenotype SW the corresponding mutation value and gene expression is 9bpins exon 2 and $g.519\_520insTGTGGTGGT$ respectively.
For the clinical phenotype Normal the corresponding mutation value and gene expression is p.A265V and $g.1851C > T$ respectively.

For the clinical phenotype SV the corresponding mutation value and gene expression is p.E320K and $g.2216G > A$ respectively.
For the clinical phenotype NC/SV the corresponding mutation value and gene expression is p.K121Q and $g.952A > C$ respectively.
For the clinical phenotype SW the corresponding mutation value and gene expression is p.T450P and $g.2786A > C$ respectively.
For the clinical phenotype SW the corresponding mutation value and gene expression is p.L142P and $g.1016T > C$ respectively.

4.1.2.8 Table 8

Figure 4.17 is reproduced from Clarke et. al [5]. In this table, the Locus actually refers to the Gene Position and a concept named Gene Position is present in the ontology. After successfully populating the ontology the following queries can be done to retrieve information.

Query VIII A: Show Gene and Locus values for Exon value(s) containing value 2.
Result: The particular Gene RHO and the corresponding Gene Location 3q21.3-q24.
The particular Gene PDE6B and the corresponding Gene Location 4p16.3.
The particular Gene RGR and the corresponding Gene Location 10q23.1.
The particular Gene RPGR and the corresponding Gene Location Xp11.4.
The particular Gene RDH12 and the corresponding Gene Location 14q23.3.
The particular Gene NR2E3 and the corresponding Gene Location 15q24.
The particular Gene CRB1 and the corresponding Gene Location 1q31-q32.1.
The particular Gene RP2 and the corresponding Gene Location Xp11.23.
The particular Gene RPE65 and the corresponding Gene Location 1p31.
The particular Gene LRAT and the corresponding Gene Location 4q31.1-q31.23.

Testing Table 7: Updated list of new CYP21A2 mutations not reported in http:www.imm.Ki.se/CYPalleles/cyp21.htm database.

Mutations	Genea	Clinical phenotype	Activity in vitro 17:OHP/progesterone	Reference
p.H38L	$g.325A > T$	-	-	Tardy 2006
p.E431K	$g.2729G > A$	-	-	Dain et al. 2006
p.F404S	$g.2552T > C$	SW	Modelling study	Baradaran-Heravi et al. 2007
p.T450P	$g.2786A > C$	SW	Modelling study	Baradaran-Heravi et al. 2007
10bpdel exon 1	$g.231_240delCTGCTGCTGC$	SV	Modelling study	Baradaran-Heravi et al. 2007
p.M1V	$g.213A > G$	CL	-	Tardy et al. 2007
p.M1L	$g.213A > C$	-	-	Tardy et al. 2007
p.Y59N	$g.387T > A$	CL	-	Tardy et al. 2007
p.Y47C	$g.352A > G$	-	-	Tardy 2007
p.W22X	$g.277G > A$	SW	N.D.	Di Pasquale et al. 2007
p.G56R	$g.378G > A$	SV	0.7 1.4	Soardi et al. 2008
p.L107R	$g.911T > G$	SW	0.4 0.3	Soardi et al. 2008
p.L107R	$g.911T > G$	SW	0.4 0.3	Soardi et al. 2008
p.L142P	$g.1016T > C$	SW	0.4 0.4	Soardi et al. 2008
p.K121Q	$g.952A > C$	NC/SV	14 19.5	Riepe et al. 2008
p.L353R	$g.2316T > G$	SW	-	Abid et al. 2008
p.R224W	$g.1558C > T$	NC	51.9 45.6	Concolino et al. 2008
p.D407N	$g.2560G > A$	NC	72.7 73.6	Concolino et al. 2008
p.A265V	$g.1851C > T$	Normal	100 92	Bleicken et al. 2009
p.W302S	$g.1962G > C$	SV	3	Bleicken et al. 2009
p.D322G	$g.2223A > G$	NC	18 27	Bleicken et al. 2009
9bpins exon 2	$g.519_520insTGTGGTGGT$	SW	Modelling study	Dubey et al. 2009
p.H119R	$g.947A > G$	NC	31.6 32.5	Concolino et al. 2009
p.I194N	$g.1368T > A$	NC	33.2 46.7	Concolino et al. 2009
p.K54X	$g.372A > T$	SW	N.D.	Concolino et al. 2009
p.N387K	$g.2502C > G$	NC	-	Wasniewska et al. 2009
p.V249A	$g.1803T > C$	Normalb	-	Concolino et al. 2009 (unpublished¿
p.L167P	$g.1199T > C$	SW	0.7 0.4	Tardy et al. 2010
p.G292D	$g.1931G > A$	SW	0.5 0.7	Tardy et al. 2010
p.E320K	$g.2216G > A$	SV	4.6 4.5	Tardy et al. 2010
p.I230T	$g.1577T > C$	NC	63.1 70.06	Tardy et al. 2010
p.R233K	$g.1586G > A$	NC/SV	15 8.1	Tardy et al. 2010
p.R369W	$g.2363C > T$	NC/SV	45.8 48.5	Tardy et al. 2010

Figure 4.16: Example of a testing Table taken from the set of new examples.

Query VIII B:

Show Locus and Protein value for particular Gene value say: USH2A

Result:

Locus : 1q41 Protein Value : Usherin.

Testing Table 8: Development of a Diagnostic Genetic Test for Simplex and Autosomal Recessive Retinitis Pigmentosa.

Gene	Locus	Protein	Exons
ABCA4	1p22.1	ATP-binding cassette, subfamily	8, 17, 19, 42, 48
CERKL	2q31.3	Ceramide kinase-like protein	5
CNGB1	16q13	Cyclic nucleotide gated channel	30
CRB1	1q31-q32.1	Crumbs (Drosophila) homolog	2, 3, 4, 5, 6, 7, 8, 9, 11, 12
CRX	19q13.3	Conerod homeobox	3, 4
LRAT	4q31.1-q31.23	Lecithin retinol acyltransferase	2
MERTK	2q14.1	c-mer proto-oncogene tyrosine kinase	11, 15, 19
NR2E3	15q24	Nuclear receptor subfamily 2, group E, member 3	2, 3, 4, 5, 6, 7, 8, 9
PDE6A	5q31.3	Phosphodiesterase 6A, cGMP-specific, rod, alpha	1, 7, 13
PDE6B	4p16.3	Phosphodiesterase 6B, cGMP-specific, rod, beta	1, 2, 3, 4, 12, 13, 14, 17, 22
RDH12	14q23.3	Retinol dehydrogenase 12	2, 3, 5, 6
RGR	10q23.1	Retinal G protein coupled receptor	2, 4, 6
RHO	3q21.3-q24	Rhodopsin	1, 2, 3, 4, 5
RLBP1	15q26	Retinaldehyde binding protein 1	6, 7
RP2	Xp11.23	Retinitis pigmentosa 2	1, 2, 4
RPE65	1p31	Retinal pigment epithelium-specific protein 65 kDa	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14
RPGR	Xp11.4	Retinitis pigmentosa GTPase regulator	2, 3, 4, 5, 6, 7, 8, 9, ORF15
TULP1	6p21.3	Tubby-like protein 1	11, 12, 13
USH2A	1q41	Usherin	3, 4, 6, 7, 9, 10, 12, 13, 43, 61

Figure 4.17: Example of a testing Table taken from the set of new examples

4.2 Conclusion

This section describes building and populating the ontology with new examples to show what situations might occur while populating the ontology with the new set of examples.

Chapter 5

Conclusions

This work reports on a table ontology that I designed to represent the tabular information in phenotype-genotype tables in scholarly biomedical papers. The populated ontology not only represents the semantics of each piece of information given by the hierarchical relations among the information but also preserves the relationship among cells in the table. After extracting tables with footnotes from the text, processing of the text from the raw data of the table is performed. Some cases where content of a cell breaks down into multiple lines, or in the case of multi-line row headings, rows need to be joined together to form a single line so that interpretation of the information and population of the ontology based on the column header is done correctly. In the case of missing column headings the information from the caption can be considered as an useful source of information to find a suitable column heading. Moreover, after processing of the text I made a copy of the tables in a format that is useful for further processing which makes the whole process more convenient. I developed a table isolation software to extract tables with multi-line footnotes for those cases where table isolation fails with the help of the *TableSeer* [19] open source software. Under a column the data can belong to two or three categories. To categorize it correctly and enter it into the ontology at the appropriate position I verified

the category with the help of other open source software to decide whether the cell refers to a phenotype, or a disease name, or a syndrome name.

The concepts of horizontal and vertical tables have been presented in order for tables to be properly interpreted. Heuristics were developed to distinguish these two types of tables. I employed the reading path concept [16] and changed the sequence of reading path in a way that make it functional for my concept of horizontal table. Furthermore, I combined all reading path in a way that make it functional for my approach of populating the proposed ontology. This approach helps to populate the ontology with data and preserve the various relationships among the table data at the same time. As the design of proposed ontology is tree structured, it should be reasonably straightforward to add new concepts when I encounter tables that contain these new concepts.

Furthermore, I built some query engines so that future querying can give informative results from the ontology. The ontology can be searched for information like mutation, corresponding gene information, position of mutation, phenotype, genotype information, etc.

5.1 Future Work

The hypothesis underlying the research discussed in this thesis was that one could take previous table research and software that came from that research and apply them to a particular domain, phenotype-genotype tables, in particular. My goal was to provide an ontology that captured how clinical and experimental data was presented in tables so that the ontology could be populated with the data from tables so that it could be used for other purposes.

What I discovered was twofold:

1. The software used to isolate the tables still has some small problems that did not allow this aspect of the project to be completely automated.

2. Hurst's work on table interpretation [16], while providing a framework for the task of table interpretation, left important aspects to be defined precisely enough to allow design and implementation of software to perform the task automatically. As well, some of the ideas presented in that work, when mapped to the phenotype-genotype table interpretation task, are not adequate, since some of the information needed in the table interpretation task is not in the table itself. Rather it is in the paper text or in the knowledge of the reader of the table.

The following are suggestions for future work that will address these two items.

1. Incorporating the multi-line footnote algorithm into *TableSeer* would improve the isolation software, making it nearly automatic.

2. New tables will generate ontological concepts not found in the current ontology. An automatic procedure for adding these concepts to the ontology could be investigated. One source of knowledge is the information found in the other parts of the paper. So a technique could be employed to find out the relevant information of a certain term from the body of the paper.

3. Some columns contain mutation values, nucleotide sequence changes, or protein sequence changes are expressed using coding techniques that can be further explained by using mutation nomenclature.

4. To facilitate table information interpretation, parsing of the footnotes can be useful. The table extraction software can retrieve multi-line footnotes. Parsing techniques can be employed to find out the useful information from the footnotes automatically.

5. To populate the ontology properly there should be a good number of synonyms for the concepts presented in the proposed ontology. In the thesis I was able to come up with some synonyms. However, the list should be expanded and built automatically.

Finally, a good representation of phenotype-genotype tables has been developed. One would obviously want to take this design and see whether it maps to other types of tables found in the scholarly biomedical literature and add the appropriate concepts when the design fails.

Bibliography

[1] Mohamad M Al-Rahawan, Deborah J Chute, Katia Sol-Church, Karen W Gripp, Deborah L Stabley, Nancy L McDaniel, William G Wilson, and Peter E Waldron. Hepatoblastoma and heart transplantation in a patient with cardio-facio-cutaneous syndrome. *Am J Med Genet A*, 143A(13):1481–8, 2007.

[2] M. A. Andrade and A. Valencia. Automatic extraction of keywords from scientific text: application to the knowledge domain of protein families. *Bioinformatics*, 14(7):600–607, 1998.

[3] G Bonuccelli, P Di Natale, F Corsolini, G Villani, S Regis, and M Filocamo. The effect of four mutations on the expression of iduronate-2-sulfatase in mucopolysaccharidosis type II. *Biochim Biophys Acta*, 1537(3):233–8, 2001.

[4] Nawei Chen, Hagit Shatkay, and Dorothea Blostein. Use of figures in literature mining for biomedical digital libraries. In *DIAL*, pages 180–197. IEEE Computer Society, 2006.

[5] Graeme Richard Clark, Paul Crowe, Dorota Muszynska, Dominic O'Prey, Justin O'Neill, Sharon Alexander, Colin Eric Willoughby, Gareth John McKay, Giuliana Silvestri, and David Arthur Simpson. Development of a diagnostic genetic test for simplex and autosomal recessive retinitis pigmentosa. *Ophthalmology*, 117(11):2169–77.e3, 2010.

[6] Paola Concolino, Enrica Mello, Cecilia Zuppi, and Ettore Capoluongo. Molecular diagnosis of congenital adrenal hyperplasia due to 21-hydroxylase deficiency: an update of new cyp21a2 mutations. *Clin Chem Lab Med*, 48(8):1057–62, 2010.

[7] C. Friedman, P. Kra, H. Yu, M. Krauthammer, and A. Rzhetsky. GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles. *Bioinformatics (Oxford, England)*, 17 Suppl 1(suppl_1):S74–82, June 2001.

[8] Katrin Fundel, Robert Küffner, and Ralf Zimmer. Relex - relation extraction using dependency parse trees. *Bioinformatics*, 23(3):365–371, 2007.

[9] R. Gaizauskas, G. Demetriou, P. J. Artymiuk, and P. Willett. Protein structures and information extraction from biological texts: the PASTA system. *Bioinformatics*, 19(1):135–143, January 2003.

[10] Xiaofen He and Chrysanne DiMarco. Using lexical chaining to rank protein-protein interactions in biomedical texts. In *BioLink 2005: Workshop on Linking Biological Literature, Ontologies and Databases: Mining Biological Semantics, Conference of the Association for Computational Linguistics (poster Presentation)*, 2005.

[11] Kristien P Hoornaert, Inge Vereecke, Chantal Dewinter, Thomas Rosenberg, Frits A Beemer, Jules G Leroy, Laila Bendix, Erik Bjòrck, Maryse Bonduelle, Odile Boute, Valerie Cormier-Daire, Christine De Die-Smulders, Anne Dieux-Coeslier, Hèlène Dollfus, Mariet Elting, Andrew Green, Veronica I Guerci, Raoul C M Hennekam, Yvonne Hilhorts-Hofstee, Muriel Holder, Carel Hoyng, Kristi J Jones, Dragana Josifova, Ilkka Kaitila, Suzanne Kjaergaard, Yolande H Kroes, Kristina Lagerstedt, Melissa Lees, Martine Lemerrer, Cinzia Magnani, Carlo Marcelis, Loreto Martorell, Michaegrave;le Mathieu, Meriel McEntagart, Angela Mendicino, Jenny Morton, Gabrielli Orazio, Vaeacute;ronique Paquis, Orit

Reish, Kalle O J Simola, Sarah F Smithson, Karen I Temple, Elisabeth Van Aken, Yolande Van Bever, Jenneke van den Ende, Johanna M Van Hagen, Leopoldo Zelante, Riina Zordania, Anne De Paepe, Bart P Leroy, Marc De Buyzere, Paul J Coucke, and Geert R Mortier. Stickler syndrome caused by col2a1 mutations: genotype-phenotype correlation in a series of 100 patients. *Eur J Hum Genet*, 18(8):872–80, 2010.

[12] D. Hristovski, C. Friedman, T. C. Rindflesch, and B. Peterlin. Exploiting semantic relations for literature-based discovery. *Proc AMIA Symp*, pages 349–353, 2006.

[13] Dimitar Hristovski, Borut Peterlin, Joyce A. Mitchell, and Susanne M. Humphrey. Using literature-based discovery to identify disease candidate genes. *I. J. Medical Informatics*, 74(2-4):289–298, 2005.

[14] K. Humphreys, G. Demetriou, and R. Gaizauskas. Two applications of information extraction to biological science journal articles: enzyme interactions and protein structures. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pages 505–516, 2000.

[15] Lawrence Hunter and K. Bretonnel Cohen. Biomedical language processing: Perspective whats beyond PubMed? *Molecular Cell*, 21(5):589–594, 2006.

[16] Matthew Francis Hurst. *The interpretation of tables in texts.* PhD thesis, The University of Edinburgh, 2000.

[17] See kiong Ng and Marie Wong. Toward routine automatic pathway discovery from on-line scientific text abstracts. *Genome Informatics*, 10:104–112, 1999.

[18] Gondy Leroy, Hsinchun Chen, and Jesse D. Martinez. A shallow parser based on closed-class words to capture relations in biomedical text. *Journal of Biomedical Informatics*, 36(3):145–158, 2003.

[19] Ying Liu. Tableseer: Automatic table extraction, search, and understanding. Master's thesis, The Pennsylvania State University, 2009.

[20] Jennifer Rowley Magdalini Vasileiou, Richard Hartley. An overview of the e-bookmarketplace. *The international journal of digital information research and use*, 33:173–192, 2008.

[21] J B Nielsen, K Ravn, and M Schwartz. A 77-year-old woman and a preserved speech variant among the danish rett patients with mutations in mecp2. *Brain Dev*, 23 Suppl 1:S230–2, 2001.

[22] G. Ovtcharov and R. Dorsch. Pdftohtml, 2008.

[23] Stavroula Psoni, Christalena Sofocleous, Joanne Traeger-Synodinos, Sophia Kitsiou-Tzeli, Emmanuel Kanavakis, and Helen Fryssira-Kanioura. Phenotypic and genotypic variability in four males with mecp2 gene sequence aberrations including a novel deletion. *Pediatr Res*, 67(5):551–6, 2010.

[24] Chloè; Quèlin, Claude Bendavid, Christèle Dubourg, Cèline de la Rochebrochard, Josette Lucas, Catherine Henry, Sylvie Jaillard, Philippe Loget, Laurence Loeuillet, Didier Lacombe, Jean-Marie Rival, Vèronique David, Sylvie Odent, and Laurent Pasquier. Twelve new patients with 13q deletion syndrome: genotype-phenotype analyses in progress. *Eur J Med Genet*, 52(1):41–6, 2009.

[25] Thomas C. Rindflesch, Lorraine Tanabe, John N. Weinstein, and Lawrence Hunter. Edgar: Extraction of drugs, genes and relations from the biomedical literature. In *Pacific Symposium on Biocomputing*, volume 5, pages 514–525, 2000.

[26] Maryam Khordad Robert E. Mercer Peter Rogan. Improving phenotype name recognition. In *C. Butz and P. Lingras (Eds.)*. Springer, Heidelberg, 2011.

[27] D.R. Swanson. Fish oil, raynauds syndrome, and undiscovered public knowledge. *Perspect. Bio. Med*, 30:7–18, 1986.

[28] L. Tanabe, U. Scherf, L. H. Smith, J. K. Lee, L. Hunter, and J. N. Weinstein. MedMiner: an Internet text-mining tool for biomedical information, with application to gene expression profiling. *BioTechniques*, 27(6), December 1999.

[29] Alfonso Valencia. Automatic annotation of protein function. *Current opinion in structural biology*, 15(3):267–274, June 2005.

[30] Wern Wong, David Martinez, and Lawrence Cavedon. Extraction of named entities from tables in gene mutation literature. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, BioNLP '09, pages 46–54, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[31] Burcu Yildiz. Information extraction: Utilizing table patterns. Master's thesis, Vienna University of Technology, 2004.

[32] Hao Yu, Xiaoyan Zhu, Minlie Huang, and Ming Li. Discovering patterns to extract protein-protein interactions from the literature: Part II. *Bioinformatics*, 21(15):3294–3300, 2005.

Appendix A

Result of Query X

A.1 Result of Query XA

The value of Growth retardation at location 13q33.1-13qter is : +

The value of Microcephaly at location 13q31.1-13qter is : +

The value of Limb defects!Club foot at location 13q32.3 is : -

The value of Microcephaly at location 13q33.1-13qter is : +

The value of Heart defects!DSCV at location 13q31.3-13q33.1 is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q21.1-13q31.1 is : -

The value of Heart defects!Aorta coarctation at location 13q13.3-13qter is : -

The value of Ocular anomalies!Coloboma at location 13q13.3-13qter is : -

The value of Facial dysmorphy!Short philtrum at location 13q33.1-13qter is : NK

The value of Facial dysmorphy!Microretrognathism at location 13q32.1-13qter is : -

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q21.32-13qter is : +

The value of Vertebral anomalies at location 13q31.1-13qter is : -

The value of Genital anomalies at location 13q33.1-13qter is : -

The value of Ocular anomalies!Micro/Anophthalmia at location 13q33.3-13qter is : -

The value of Mental retardation at location 13q13.3-13qter is : NK

The value of Genital anomalies at location 13q31.3-13q34 is : +

The value of Ocular anomalies!Coloboma at location 13q31.1-13q33.3 is : -

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q21.1-13q31.1 is : NK

The value of Heart defects!Aorta coarctation at location 13q31.3-13q34 is : -

The value of Limb defects!Single transverse palmar crease at location 13q13.3-13qter is : Unilateral A The value of Limb defects!Clinodactyly V at location 13q33.1-13qter is : -

The value of Limb defects!Club foot at location 13q31.3-13q33.1 is : -

The value of Displaced anus/anal atresia at location 13q31.3-13q34 is : -

The value of Heart defects!Interauricular communication at location 13q31.3-13q33.1 is : -

The value of Microcephaly at location 13q31.3-13q34 is : +

The value of Limb defects!Club foot at location 13q31.1-13q33.3 is : +

The value of Brain anomalies!Corpus callosum agenesis at location 13q32.1-13qter is : NK

The value of Ocular anomalies!Retinal dysplasia at location 13q33.1-13qter is : -

The value of Microcephaly at location 13q31.3-13q33.1 is : -

The value of Heart defects!Tetralogy of Fallot at location 13q32.1-13qter is : -

The value of Vertebral anomalies at location 13q31.3-13q34 is : -

The value of Digestive anomalies!Spleen at location 13q31.3-13q33.1 is : -

The value of Genital anomalies at location 13q31.1-13qter is : +

The value of Heart defects!Aorta coarctation at location 13q31.1-13qter is : -

The value of Facial dysmorphy!Hypotelorism at location 13q33.3-13qter is : NK

The value of Limb defects!Club foot at location 13q13.3-13qter is : -

The value of Cleft lip and palate at location 13q31.1-13qter is : -

The value of Digestive anomalies!Pancreas anomalies at location 13q31.1-13qter is : -

The value of Facial dysmorphy!Thick and short neck at location 13q31.1-13q33.3 is : NK

The value of Digestive anomalies!Esophageal atresia at location 13q21.32-13qter is : -

The value of Facial dysmorphy!Thick and short neck at location 13q31.3-13q34 is : +

The value of Brain anomalies!Cortical dysplasia at location 13q31.1-13qter is : -

The value of Ocular anomalies!Cataract at location 13q31.3-13q33.1 is : -

The value of Heart defects!Aorta coarctation at location 13q32.1-13qter is : +

The value of Limb defects!Camptodactyly at location 13q32.1-13qter is : -

The value of Ocular anomalies!Micro/Anophthalmia at location 13q31.3-13q33.1 is : -

The value of Heart defects!DSCV at location 13q33.1-13qter is : -

The value of Brain anomalies!Aprosencephaly at location 13q31.1-13qter is : -

The value of Ocular anomalies!Retinal dysplasia at location 13q31.1-13qter is : +

The value of Mental retardation at location 13q21.1-13q31.1 is : +

The value of Ocular anomalies!Blepharophimosis at location 13q21.32-13qter is : -

The value of Brain anomalies!Aprosencephaly at location 13q32.3 is : -

The value of Limb defects!Thumb agenesis at location 13q13.3-13qter is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q33.1-13qter is : -

The value of Digestive anomalies!Spleen at location 13q33.1-13qter is : -

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q31.1-13qter is : +

The value of Heart defects!ARSA at location 13q31.3-13q33.1 is : -

The value of Facial dysmorphy!Hypertelorism at location 13q13.3-13qter is : +

The value of Growth retardation at location 13q31.1-13q33.3 is : NK

The value of Displaced anus/anal atresia at location 13q21.32-13qter is : +

The value of Facial dysmorphy!Thick and short neck at location 13q32.1-13qter is : +

The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q21.1-13q31.1 is : -

The value of Limb defects!Camptodactyly at location 13q21.32-13qter is : -

The value of Limb defects!Club foot at location 13q33.1-13qter is : -

The value of Digestive anomalies!Pancreas anomalies at location 13q32.1-13qter is : -

The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q31.1-13qter is : -

The value of Displaced anus/anal atresia at location 13q33.1-13qter is : NK

The value of Facial dysmorphy!Microretrognathism at location 13q33.1-13qter is : +

The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q33.1-13qter is : -

The value of Mental retardation at location 13q31.1-13q33.3 is : NK

The value of Heart defects!ARSA at location 13q13.3-13qter is : -

The value of Limb defects!Clinodactyly V at location 13q31.3-13q33.1 is : +

The value of Growth retardation at location 13q21.1-13q31.1 is : +

The value of Facial dysmorphy!Short philtrum at location 13q32.1 -13qter is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q31.3-13q33.1 is : NK

The value of Heart defects!Aorta coarctation at location 13q33.1-13qter is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q33.3-13qter is : -

The value of Growth retardation at location 13q13.3-13qter is : NK

The value of Facial dysmorphy!Hypotelorism at location 13q32.3 is : NK

The value of Growth retardation at location 13q31.3-13q33.1 is : NK

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q31.3-13q33.1 is : +

The value of Limb defects!Clinodactyly V at location 13q33.3-13qter is : -

The value of Digestive anomalies!Pancreas anomalies at location 13q21.32-13qter is : -

The value of Ocular anomalies!Coloboma at location 13q32.3 is : -

The value of Digestive anomalies!Esophageal atresia at location 13q13.3-13qter is : -

The value of Facial dysmorphy!Thick and short neck at location 13q32.3 is : NK

The value of Displaced anus/anal atresia at location 13q32.1-13qter is : -

The value of Heart defects!Interauricular communication at location 13q21.1-13q31.1 is : -

The value of Limb defects!Camptodactyly at location 13q31.1-13q33.3 is : -

The value of Brain anomalies!Cortical dysplasia at location 13q21.32-13qter is : -

The value of Heart defects!Interauricular communication at location 13q31.3-13q34 is : -

The value of Heart defects!Tetralogy of Fallot at location 13q21.32-13qter is : -

The value of Digestive anomalies!Esophageal atresia at location 13q31.3-13q33.1 is : -

The value of Limb defects!Club foot at location 13q31.1-13qter is : +

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q32.3 is : +

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q33.3-13qter is : NK

The value of Heart defects!Tetralogy of Fallot at location 13q21.1-13q31.1 is : -

The value of Facial dysmorphy!Microretrognathism at location 13q31.1-13qter is : NK

The value of Facial dysmorphy!Hypertelorism at location 13q33.1-13qter is : +

The value of IUGR at location 13q31.1-13q33.3 is : -

The value of Limb defects!Camptodactyly at location 13q31.3-13q33.1 is : -

The value of Ocular anomalies!Cataract at location 13q32.1-13qter is : NK

The value of Digestive anomalies!Spleen at location 13q31.1-13q33.3 is : NK

The value of IUGR at location 13q31.3-13q34 is : -

The value of Digestive anomalies!Spleen at location 13q31.3-13q34 is : -

The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q33.1-13qter is : -

The value of Brain anomalies!Holoprosencephaly at location 13q31.1-13qter is : +

The value of Limb defects!Single transverse palmar crease at location 13q31.1-13qter is : -

The value of Facial dysmorphy!Malformed ears at location 13q31.3-13q34 is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q32.3 is : NK

The value of IUGR at location 13q33.1-13qter is : +

The value of Brain anomalies!Holoprosencephaly at location 13q32.3 is : +

The value of Limb defects!Clinodactyly V at location 13q21.32-13qter is : -

The value of Mental retardation at location 13q31.3-13q34 is : NK

The value of Cleft lip and palate at location 13q13.3-13qter is : -

The value of Cleft lip and palate at location 13q31.3-13q33.1 is : -

The value of Digestive anomalies!Pancreas anomalies at location 13q31.1-13qter is : -

The value of Heart defects!DSCV at location 13q13.3-13qter is : -

The value of Ocular anomalies!Blepharophimosis at location 13q33.3-13qter is : -

The value of Brain anomalies!Cortical dysplasia at location 13q32.1-13qter is : NK

The value of Digestive anomalies!Esophageal atresia at location 13q31.1-13q33.3 is : -

The value of Heart defects!Interauricular communication at location 13q21.32-13qter is : -

The value of Microcephaly at location 13q21.1-13q31.1 is : -

The value of Facial dysmorphy!Hypertelorism at location 13q32.3 is : NK

The value of Heart defects!Interauricular communication at location 13q31.1-13qter is : -

The value of Cleft lip and palate at location 13q33.3-13qter is : -

The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q31.1-13qter is : -

The value of Heart defects!DSCV at location 13q31.1-13q33.3 is : -

The value of Facial dysmorphy!Short philtrum at location 13q21.1-13q31.1 is : NK

The value of Digestive anomalies!Pancreas anomalies at location 13q31.1-13q33.3 is : +

The value of Digestive anomalies!Gall bladder agenesis at location 13q13.3-13qter is : -

The value of Ocular anomalies!Retinal dysplasia at location 13q31.3-13q34 is : -

The value of Brain anomalies!Aprosencephaly at location 13q21.32-13qter is : -

The value of Digestive anomalies!Esophageal atresia at location 13q33.1-13qter is : -

The value of Genital anomalies at location 13q31.1-13q33.3 is : -

The value of Heart defects!Interauricular communication at location 13q32.1-13qter is : +

The value of Facial dysmorphy!Hypotelorism at location 13q21.32-13qter is : NK

The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q32.1-13qter is : -

The value of Facial dysmorphy!Malformed ears at location 13q21.32-13qter is : +

The value of Digestive anomalies!Gall bladder agenesis at location 13q32.3 is : -

The value of Brain anomalies!Holoprosencephaly at location 13q13.3-13qter is : -

The value of Heart defects!Tetralogy of Fallot at location 13q33.3-13qter is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q31.3-13q34 is : NK

The value of Cleft lip and palate at location 13q31.1-13q33.3 is : -

The value of Digestive anomalies!Spleen at location 13q33.3-13qter is : -

The value of Ocular anomalies!Coloboma at location 13q33.1-13qter is : -

The value of Facial dysmorphy!Hypertelorism at location 13q21.1-13q31.1 is : +

The value of Mental retardation at location 13q33.3-13qter is : +

The value of Ocular anomalies!Cataract at location 13q13.3-13qter is : -

The value of Limb defects!Thumb agenesis at location 13q31.1-13qter is : -

The value of Facial dysmorphy!Microretrognathism at location 13q13.3-13qter is : NK

The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q21.1-13q31.1 is : -

The value of Ocular anomalies!Retinal dysplasia at location 13q21.1-13q31.1 is : -

The value of Growth retardation at location 13q32.3 is : NK

The value of Digestive anomalies!Pancreas anomalies at location 13q32.3 is : -

The value of Facial dysmorphy!Microretrognathism at location 13q31.1-13q33.3 is : +

The value of Heart defects!DSCV at location 13q31.1-13qter is : -

The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q31.3-13q33.1 is : NK

The value of Displaced anus/anal atresia at location 13q21.1-13q31.1 is : NK

The value of Ocular anomalies!Micro/Anophthalmia at location 13q32.3 is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q31.1-13qter is : -

The value of Facial dysmorphy!Malformed ears at location 13q32.3 is : NK

The value of Ocular anomalies!Blepharophimosis at location 13q31.3-13q33.1 is : -

The value of Facial dysmorphy!Short philtrum at location 13q33.3-13qter is : NK

The value of Facial dysmorphy!Hypotelorism at location 13q31.1-13qter is : +

The value of Facial dysmorphy!Hypotelorism at location 13q31.3-13q33.1 is : NK

The value of Limb defects!Clinodactyly V at location 13q32.1-13qter is : -

The value of Heart defects!ARSA at location 13q33.1-13qter is : -

The value of Ocular anomalies!Coloboma at location 13q31.1-13qter is : +

The value of Digestive anomalies!Esophageal atresia at location 13q21.1-13q31.1 is : -

The value of Facial dysmorphy!Thick and short neck at location 13q33.1-13qter is :

NK

The value of Limb defects!Camptodactyly at location 13q31.1-13qter is : +

The value of Brain anomalies!Cortical dysplasia at location 13q31.3-13q34 is : NK

The value of Ocular anomalies!Retinal dysplasia at location 13q21.32-13qter is : -

The value of Facial dysmorphy!Hypertelorism at location 13q31.3-13q34 is : NK

The value of Brain anomalies!Corpus callosum agenesis at location 13q31.1-13q33.3 is : -

The value of Limb defects!Camptodactyly at location 13q32.3 is : -

The value of Facial dysmorphy!Malformed ears at location 13q21.1-13q31.1 is : NK

The value of Ocular anomalies!Micro/Anophthalmia at location 13q31.1-13q33.3 is : -

The value of IUGR at location 13q31.1-13qter is : +

The value of Heart defects!Interauricular communication at location 13q33.3-13qter is : -

The value of Facial dysmorphy!Short philtrum at location 13q31.1-13qter is : +

The value of Heart defects!Tetralogy of Fallot at location 13q31.3-13q33.1 is : -

The value of Genital anomalies at location 13q21.32-13qter is : +

The value of Brain anomalies!Aprosencephaly at location 13q31.3-13q33.1 is : -

The value of Facial dysmorphy!Microretrognathism at location 13q31.3-13q34 is : +

The value of Limb defects!Single transverse palmar crease at location 13q31.3-13q33.1 is : -

The value of Limb defects!Thumb agenesis at location 13q21.1-13q31.1 is : -

The value of Facial dysmorphy!Microretrognathism at location 13q32.3 is : NK

The value of Heart defects!DSCV at location 13q31.1-13qter is : +

The value of Heart defects!Tetralogy of Fallot at location 13q32.3 is : -

The value of IUGR at location 13q33.3-13qter is : +

The value of Limb defects!Single transverse palmar crease at location 13q33.1-13qter is : -

The value of Genital anomalies at location 13q31.1-13qter is : +

The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q21.32-13qter is : +

The value of Limb defects!Single transverse palmar crease at location 13q32.3 is : -

The value of Facial dysmorphy!Thick and short neck at location 13q21.1-13q31.1 is : NK

The value of Limb defects!Thumb agenesis at location 13q33.1-13qter is : -

The value of Ocular anomalies!Cataract at location 13q31.1-13qter is : -

The value of Heart defects!Interauricular communication at location 13q31.1-13qter is : -

The value of Ocular anomalies!Micro/Anophthalmia at location 13q13.3-13qter is : -

The value of Limb defects!Camptodactyly at location 13q13.3-13qter is : -

The value of Facial dysmorphy!Short philtrum at location 13q21.32-13qter is : +

The value of Ocular anomalies!Cataract at location 13q31.1-13q33.3 is : -

The value of Digestive anomalies!Esophageal atresia at location 13q31.3-13q34 is : Fistula The value of Digestive anomalies!Gall bladder agenesis at location 13q31.1-13q33.3 is : -

The value of Brain anomalies!Corpus callosum agenesis at location 13q31.1-13qter is : NK

The value of Digestive anomalies!Spleen at location 13q21.1-13q31.1 is : -

The value of Heart defects!DSCV at location 13q32.1-13qter is : -

The value of Brain anomalies!Aprosencephaly at location 13q31.1-13q33.3 is : -

The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q31.1-13qter is : +

The value of Microcephaly at location 13q31.1-13qter is : +

The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q33.3-13qter is : -

The value of Facial dysmorphy!Hypotelorism at location 13q31.1-13q33.3 is : NK

The value of Microcephaly at location 13q33.3-13qter is : -

The value of Facial dysmorphy!Malformed ears at location 13q31.1-13qter is : +

The value of Heart defects!DSCV at location 13q31.3-13q34 is : -

The value of Genital anomalies at location 13q32.1-13qter is : -

The value of Ocular anomalies!Micro/Anophthalmia at location 13q31.3-13q34 is : -

The value of Digestive anomalies!Gall bladder agenesis at location 13q31.3-13q33.1 is : -

The value of Brain anomalies!Holoprosencephaly at location 13q31.1-13q33.3 is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q21.32-13qter is : -

The value of Heart defects!Aorta coarctation at location 13q21.1-13q31.1 is : -

The value of Vertebral anomalies at location 13q21.1-13q31.1 is : NK

The value of Vertebral anomalies at location 13q13.3-13qter is : +

The value of Genital anomalies at location 13q33.3-13qter is : -

The value of Limb defects!Thumb agenesis at location 13q31.3-13q34 is : -

The value of Ocular anomalies!Blepharophimosis at location 13q32.1-13qter is : NK

The value of Brain anomalies!Holoprosencephaly at location 13q33.1-13qter is : -

The value of Heart defects!ARSA at location 13q32.3 is : -

The value of Growth retardation at location 13q33.3-13qter is : +

The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q32.3 is : NK

The value of Facial dysmorphy!Microretrognathism at location 13q31.3-13q33.1 is : NK

The value of Facial dysmorphy!Hypertelorism at location 13q33.3-13qter is : +

The value of Digestive anomalies!Pancreas anomalies at location 13q31.3-13q34 is : +

The value of Limb defects!Club foot at location 13q31.3-13q34 is : -

The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q21.32-

13qter is : -

The value of Microcephaly at location 13q32.1-13qter is : +

The value of Brain anomalies!Aprosencephaly at location 13q32.1-13qter is : +

? The value of Ocular anomalies!Retinal dysplasia at location 13q31.1-13qter is : -

The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q33.3-13qter is : -

The value of Brain anomalies!Aprosencephaly at location 13q13.3-13qter is : -

The value of Vertebral anomalies at location 13q31.1-13qter is : -

The value of Ocular anomalies!Retinal dysplasia at location 13q33.3-13qter is : -

The value of Facial dysmorphy!Malformed ears at location 13q32.1-13qter is : +

The value of Displaced anus/anal atresia at location 13q33.3-13qter is : NK

The value of Vertebral anomalies at location 13q32.1-13qter is : -

The value of Heart defects!ARSA at location 13q31.1-13q33.3 is : -

The value of Limb defects!Clinodactyly V at location 13q31.1-13q33.3 is : +

The value of Limb defects!Clinodactyly V at location 13q32.3 is : -

The value of Digestive anomalies!Pancreas anomalies at location 13q13.3-13qter is : -

The value of Ocular anomalies!Coloboma at location 13q31.3-13q34 is : -

The value of Digestive anomalies!Gall bladder agenesis at location 13q33.1-13qter is : -

The value of Limb defects!Club foot at location 13q21.1-13q31.1 is : -

The value of Brain anomalies!Cortical dysplasia at location 13q31.1-13qter is : NK

The value of Limb defects!Thumb agenesis at location 13q32.1-13qter is : -

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q31.3-13q34 is : +

The value of Facial dysmorphy!Hypertelorism at location 13q21.32-13qter is : +

The value of Brain anomalies!Corpus callosum agenesis at location 13q13.3-13qter is : -

The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q31.3-13q34

is : NK

The value of Displaced anus/anal atresia at location 13q13.3-13qter is : +

The value of Digestive anomalies!Gall bladder agenesis at location 13q21.1-13q31.1 is : -

The value of Ocular anomalies!Cataract at location 13q32.3 is : -

The value of IUGR at location 13q31.1-13qter is : +

The value of Digestive anomalies!Esophageal atresia at location 13q33.3-13qter is : -

The value of Facial dysmorphy!Malformed ears at location 13q31.1-13q33.3 is : +

The value of Brain anomalies!Holoprosencephaly at location 13q21.1-13q31.1 is : -

The value of Ocular anomalies!Coloboma at location 13q21.1-13q31.1 is : -

The value of IUGR at location 13q32.1-13qter is : +

The value of Vertebral anomalies at location 13q32.3 is : -

The value of Ocular anomalies!Blepharophimosis at location 13q31.1-13qter is : -

The value of Digestive anomalies!Spleen at location 13q32.1-13qter is : -

The value of Facial dysmorphy!Thick and short neck at location 13q21.32-13qter is : +

The value of Facial dysmorphy!Thick and short neck at location 13q31.1-13qter is : NK

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q31.1-13q33.3 is : NK

The value of Brain anomalies!Aprosencephaly at location 13q33.1-13qter is : -

The value of Displaced anus/anal atresia at location 13q31.1-13qter is : -

The value of Limb defects!Thumb agenesis at location 13q31.1-13qter is : +

The value of Limb defects!Thumb agenesis at location 13q33.3-13qter is : -

The value of Cleft lip and palate at location 13q32.3 is : -

The value of Growth retardation at location 13q31.1-13qter is : NK

The value of Limb defects!Camptodactyly at location 13q21.1-13q31.1 is : -

The value of Heart defects!ARSA at location 13q31.1-13qter is : +

The value of Heart defects!DSCV at location 13q21.1-13q31.1 is : -

The value of Digestive anomalies!Spleen at location 13q21.32-13qter is : Hypoplasia

A The value of Facial dysmorphy!Thick and short neck at location 13q33.3-13qter is : NK

The value of Microcephaly at location 13q21.32-13qter is : -

The value of Genital anomalies at location 13q21.1-13q31.1 is : -

The value of Limb defects!Clinodactyly V at location 13q13.3-13qter is : +

The value of Limb defects!Club foot at location 13q21.32-13qter is : +

The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q31.3-13q33.1 is : -

The value of Limb defects!Club foot at location 13q32.1-13qter is : +

The value of Vertebral anomalies at location 13q31.1-13q33.3 is : -

The value of Ocular anomalies!Micro/Anophthalmia at location 13q33.1-13qter is : -

The value of Facial dysmorphy!Malformed ears at location 13q33.1-13qter is : NK

The value of Heart defects!Aorta coarctation at location 13q21.32-13qter is : -

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q13.3-13qter is : NK

The value of Facial dysmorphy!Hypotelorism at location 13q31.1-13qter is : +

The value of Heart defects!Aorta coarctation at location 13q31.3-13q33.1 is : -

The value of Facial dysmorphy!Hypotelorism at location 13q31.3-13q34 is : -

The value of Mental retardation at location 13q31.1-13qter is : NK

The value of Limb defects!Clinodactyly V at location 13q31.3-13q34 is : -

The value of Heart defects!Aorta coarctation at location 13q31.1-13qter is : -

The value of Brain anomalies!Cortical dysplasia at location 13q31.3-13q33.1 is : +

The value of Mental retardation at location 13q32.1-13qter is : NK

The value of Brain anomalies!Holoprosencephaly at location 13q33.3-13qter is : -

The value of Brain anomalies!Cortical dysplasia at location 13q13.3-13qter is : -

The value of Heart defects!Aorta coarctation at location 13q33.3-13qter is : -

The value of Ocular anomalies!Coloboma at location 13q33.3-13qter is : -
The value of Brain anomalies!Cortical dysplasia at location 13q32.3 is : ? The value of Growth retardation at location 13q31.3-13q34 is : NK
The value of Facial dysmorphy!Hypertelorism at location 13q32.1-13qter is : -
The value of Vertebral anomalies at location 13q33.3-13qter is : NK
The value of Ocular anomalies!Cataract at location 13q21.1-13q31.1 is : +
The value of Brain anomalies!Corpus callosum agenesis at location 13q31.1-13qter is : +
The value of Displaced anus/anal atresia at location 13q32.3 is : -
The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q31.1-13q33.3 is : -
The value of Heart defects!Tetralogy of Fallot at location 13q31.1-13q33.3 is : -
The value of Brain anomalies!Corpus callosum agenesis at location 13q32.3 is : NK
The value of Brain anomalies!Aprosencephaly at location 13q21.1-13q31.1 is : -
The value of Ocular anomalies!Blepharophimosis at location 13q13.3-13qter is : -
The value of Brain anomalies!Aprosencephaly at location 13q31.3-13q34 is : +
The value of Digestive anomalies!Spleen at location 13q31.1-13qter is : NK
The value of Limb defects!Single transverse palmar crease at location 13q31.1-13q33.3 is : -
The value of Limb defects!Thumb agenesis at location 13q21.32-13qter is : +
The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q31.1-13qter is : NK
The value of Digestive anomalies!Gall bladder agenesis at location 13q31.1-13qter is : +
The value of Ocular anomalies!Blepharophimosis at location 13q31.1-13q33.3 is : -
The value of Brain anomalies!Holoprosencephaly at location 13q31.3-13q34 is : NK
The value of Facial dysmorphy!Thick and short neck at location 13q13.3-13qter is : +

The value of Mental retardation at location 13q31.1-13qter is : NK

The value of Digestive anomalies!Pancreas anomalies at location 13q21.1-13q31.1 is : -

The value of Ocular anomalies!Blepharophimosis at location 13q32.3 is : -

The value of Brain anomalies!Cortical dysplasia at location 13q31.1-13q33.3 is : -

The value of Digestive anomalies!Gall bladder agenesis at location 13q33.3-13qter is : -

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q33.1-13qter is : +

The value of Limb defects!Thumb agenesis at location 13q32.3 is : -

The value of Heart defects!ARSA at location 13q31.3-13q34 is : -

The value of Facial dysmorphy!Hypertelorism at location 13q31.1-13q33.3 is : +

The value of Digestive anomalies!Pancreas anomalies at location 13q31.3-13q33.1 is : -

The value of Cleft lip and palate at location 13q31.3-13q34 is : -

The value of Displaced anus/anal atresia at location 13q31.1-13q33.3 is : +

The value of Brain anomalies!Corpus callosum agenesis at location 13q21.1-13q31.1 is : -

The value of Heart defects!Tetralogy of Fallot at location 13q13.3-13qter is : -

The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q32.1-13qter is : NK

The value of Limb defects!Camptodactyly at location 13q31.3-13q34 is : -

The value of Ocular anomalies!Cataract at location 13q33.1-13qter is : -

The value of Heart defects!Interauricular communication at location 13q31.1-13q33.3 is : -

The value of Facial dysmorphy!Hypotelorism at location 13q13.3-13qter is : NK

The value of Limb defects!Club foot at location 13q33.3-13qter is : -

The value of Ocular anomalies!Micro/Anophthalmia at location 13q31.1-13qter is :

+

The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q31.1-13q33.3 is : +

The value of Ocular anomalies!Coloboma at location 13q21.32-13qter is : -

The value of Heart defects!Tetralogy of Fallot at location 13q31.1-13qter is : +

The value of Cleft lip and palate at location 13q21.1-13q31.1 is : -

The value of Digestive anomalies!Spleen at location 13q32.3 is : -

The value of Facial dysmorphy!Thick and short neck at location 13q31.1-13qter is : NK

The value of Limb defects!Single transverse palmar crease at location 13q31.3-13q34 is : -

The value of Displaced anus/anal atresia at location 13q31.3-13q33.1 is : -

The value of Ocular anomalies!Coloboma at location 13q31.1-13qter is : -

The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q13.3-13qter is : +

The value of Heart defects!Aorta coarctation at location 13q32.3 is : -

The value of Cleft lip and palate at location 13q33.1-13qter is : -

The value of Growth retardation at location 13q31.1-13qter is : NK

The value of Facial dysmorphy!Hypertelorism at location 13q31.1-13qter is : NK

The value of Facial dysmorphy!Short philtrum at location 13q13.3-13qter is : NK

The value of Facial dysmorphy!Microretrognathism at location 13q21.1-13q31.1 is : NK

The value of Microcephaly at location 13q31.1-13q33.3 is : -

The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q32.3 is : -

The value of Ocular anomalies!Micro/Anophthalmia at location 13q21.1-13q31.1 is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q32.1-13qter

is : NK

The value of IUGR at location 13q21.1-13q31.1 is : NK

The value of Limb defects!Single transverse palmar crease at location 13q21.1-13q31.1 is : -

The value of Facial dysmorphy!Short philtrum at location 13q31.3-13q34 is : -

The value of Brain anomalies!Holoprosencephaly at location 13q31.1-13qter is : +

The value of Ocular anomalies!Micro/Anophthalmia at location 13q32.1-13qter is : +

The value of Genital anomalies at location 13q13.3-13qter is : +

The value of Genital anomalies at location 13q32.3 is : -

The value of Limb defects!Clinodactyly V at location 13q31.1-13qter is : -

The value of Mental retardation at location 13q33.1-13qter is : +

The value of Facial dysmorphy!Thick and short neck at location 13q31.3-13q33.1 is : +

The value of Ocular anomalies!Blepharophimosis at location 13q31.3-13q34 is : -

The value of Mental retardation at location 13q32.3 is : NK

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q32.1-13qter is : +

The value of Facial dysmorphy!Hypertelorism at location 13q31.1-13qter is : NK

The value of Facial dysmorphy!Microretrognathism at location 13q21.32-13qter is : +

The value of Facial dysmorphy!Microretrognathism at location 13q31.1-13qter is : NK

The value of Digestive anomalies!Esophageal atresia at location 13q31.1-13qter is : -

The value of Heart defects!ARSA at location 13q33.3-13qter is : -

The value of Limb defects!Camptodactyly at location 13q31.1-13qter is : -

The value of Digestive anomalies!Pancreas anomalies at location 13q33.1-13qter is : -

The value of Ocular anomalies!Retinal dysplasia at location 13q31.1-13q33.3 is : -

The value of Brain anomalies!Cortical dysplasia at location 13q33.1-13qter is : -
The value of Ocular anomalies!Retinal dysplasia at location 13q32.1-13qter is : NK
The value of Limb defects!Camptodactyly at location 13q33.1-13qter is : -
The value of Limb defects!Single transverse palmar crease at location 13q31.1-13qter is : -
The value of Brain anomalies!Corpus callosum agenesis at location 13q33.1-13qter is : -
The value of Heart defects!Tetralogy of Fallot at location 13q31.3-13q34 is : -
The value of Limb defects!Clinodactyly V at location 13q21.1-13q31.1 is : -
The value of Vertebral anomalies at location 13q31.3-13q33.1 is : -
The value of Heart defects!ARSA at location 13q21.1-13q31.1 is : -
The value of Ocular anomalies!Blepharophimosis at location 13q31.1-13qter is : -
The value of Brain anomalies!Aprosencephaly at location 13q33.3-13qter is : -
The value of Facial dysmorphy!Hypertelorism at location 13q31.3-13q33.1 is : Epican-Pus A The value of Limb defects!Single transverse palmar crease at location 13q33.3-13qter is : -
The value of Microcephaly at location 13q13.3-13qter is : +
The value of Heart defects!ARSA at location 13q32.1-13qter is : -
The value of Cleft lip and palate at location 13q32.1-13qter is : -
The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q31.1-13q33.3 is : -
The value of Heart defects!DSCV at location 13q21.32-13qter is : -
The value of Digestive anomalies!Pancreas anomalies at location 13q33.3-13qter is : -
The value of Ocular anomalies!Cataract at location 13q31.3-13q34 is : -
The value of Facial dysmorphy!Short philtrum at location 13q31.1-13qter is : +
The value of Facial dysmorphy!Short philtrum at location 13q31.3-13q33.1 is : NK
The value of Limb defects!Metacarpal synostosis/ syndactyly at location 13q31.3-13q34 is : -

The value of IUGR at location 13q21.32-13qter is : +

The value of Limb defects!Camptodactyly at location 13q33.3-13qter is : -

The value of Digestive anomalies!Gall bladder agenesis at location 13q31.1-13qter is : -

The value of Ocular anomalies!Micro/Anophthalmia at location 13q31.1-13qter is : -

The value of Heart defects!Tetralogy of Fallot at location 13q33.1-13qter is : -

The value of Brain anomalies!Corpus callosum agenesis at location 13q33.3-13qter is : -

The value of Heart defects!DSCV at location 13q33.3-13qter is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia at location 13q13.3-13qter is : -

The value of Facial dysmorphy!Hypotelorism at location 13q32.1-13qter is : -

The value of Cleft lip and palate at location 13q21.32-13qter is : +

The value of Limb defects!Single transverse palmar crease at location 13q32.1-13qter is : -

The value of Limb defects!Thumb agenesis at location 13q31.1-13q33.3 is : -

The value of Ocular anomalies!Coloboma at location 13q31.3-13q33.1 is : -

The value of Brain anomalies!Cortical dysplasia at location 13q21.1-13q31.1 is : -

The value of Ocular anomalies!Retinal dysplasia at location 13q13.3-13qter is : -

The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q13.3-13qter is : +

The value of Growth retardation at location 13q32.1-13qter is : NK

The value of Ocular anomalies!Blepharophimosis at location 13q33.1-13qter is : +

The value of Facial dysmorphy!Short philtrum at location 13q31.1-13q33.3 is : NK

The value of Digestive anomalies!Spleen at location 13q31.1-13qter is : Supernumerary A The value of Brain anomalies!Cerebellar vermis hypoplasia at location 13q31.1-13qter is : NK

The value of Ocular anomalies!Coloboma at location 13q32.1-13qter is : NK

The value of Heart defects!Interauricular communication at location 13q32.3 is : -

The value of Digestive anomalies!Esophageal atresia at location 13q32.1-13qter is : -

The value of Brain anomalies!Corpus callosum agenesis at location 13q31.3-13q33.1 is : NK

The value of IUGR at location 13q31.3-13q33.1 is : -

The value of Brain anomalies!Corpus callosum agenesis at location 13q31.3-13q34 is : NK

The value of Ocular anomalies!Retinal dysplasia at location 13q32.3 is : -

The value of Facial dysmorphy!Malformed ears at location 13q13.3-13qter is : +

The value of Limb defects!Single transverse palmar crease at location 13q21.32-13qter is : -

The value of Microcephaly at location 13q32.3 is : -

The value of Ocular anomalies!Cataract at location 13q33.3-13qter is : -

The value of Facial dysmorphy!Microretrognathism at location 13q33.3-13qter is : NK

The value of IUGR at location 13q32.3 is : -

The value of Brain anomalies!Holoprosencephaly at location 13q31.3-13q33.1 is : +

The value of Brain anomalies!Holoprosencephaly at location 13q32.1-13qter is : NK

The value of Mental retardation at location 13q21.32-13qter is : NK

The value of Facial dysmorphy!Malformed ears at location 13q33.3-13qter is : NK

The value of Limb defects!Clinodactyly V at location 13q31.1-13qter is : -

The value of Heart defects!Aorta coarctation at location 13q31.1-13q33.3 is : -

The value of Facial dysmorphy!Hypotelorism at location 13q33.1-13qter is : NK

The value of Cleft lip and palate at location 13q31.1-13qter is : +

The value of Digestive anomalies!Esophageal atresia at location 13q31.1-13qter is : -

The value of Ocular anomalies!Micro/Anophthalmia at location 13q21.32-13qter is : -

The value of Brain anomalies!Corpus callosum agenesis at location 13q21.32-13qter

is : +

The value of Displaced anus/anal atresia at location 13q31.1-13qter is : +

The value of Ocular anomalies!Retinal dysplasia at location 13q31.3-13q33.1 is : -

The value of Ocular anomalies!Cataract at location 13q31.1-13qter is : -

The value of Facial dysmorphy!Hypotelorism at location 13q21.1-13q31.1 is : NK

The value of Digestive anomalies!Esophageal atresia at location 13q32.3 is : -

The value of Brain anomalies!Cortical dysplasia at location 13q33.3-13qter is : -

The value of Genital anomalies at location 13q31.3-13q33.1 is : -

The value of Facial dysmorphy!Flat and broad nasal bridge at location 13q31.1-13qter is : +

(1) The value of IUGR at location 13q13.3-13qter is : +

The value of Digestive anomalies!Gall bladder agenesis at location 13q31.3-13q34 is : Hypoplasia A The value of Digestive anomalies!Gall bladder agenesis at location 13q21.32-13qter is : -

The value of Heart defects!DSCV at location 13q32.3 is : -

The value of Brain anomalies!Aprosencephaly at location 13q31.1-13qter is : -

The value of Digestive anomalies!Spleen at location 13q13.3-13qter is : -

The value of Facial dysmorphy!Short philtrum at location 13q32.3 is : NK

The value of Digestive anomalies!Gall bladder agenesis at location 13q32.1-13qter is : NK

The value of Heart defects!Interauricular communication at location 13q33.1-13qter is : +

The value of Heart defects!Tetralogy of Fallot at location 13q31.1-13qter is : +

The value of Heart defects!ARSA at location 13q21.32-13qter is : -

The value of Ocular anomalies!Blepharophimosis at location 13q21.1-13q31.1 is : -

The value of Facial dysmorphy!Malformed ears at location 13q31.1-13qter is : +

The value of Facial dysmorphy!Malformed ears at location 13q31.3-13q33.1 is : NK

The value of Vertebral anomalies at location 13q21.32-13qter is : +

The value of Brain anomalies!Holoprosencephaly at location 13q21.32-13qter is : -

The value of Vertebral anomalies at location 13q33.1-13qter is : -

The value of Heart defects!ARSA at location 13q31.1-13qter is : -

The value of Heart defects!Interauricular communication at location 13q13.3-13qter is : NK

The value of Mental retardation at location 13q31.3-13q33.1 is : NK

The value of Growth retardation at location 13q21.32-13qter is : NK

The value of Ocular anomalies!Cataract at location 13q21.32-13qter is : -

The value of Limb defects!Thumb agenesis at location 13q31.3-13q33.1 is : -

The value of Limb defects!Club foot at location 13q31.1-13qter is : -

A.2 Result of Query XB

The value of Cleft lip and palate for the value of size of deletion 30 Mb is : -

The value of Digestive anomalies!Spleen for the value of size of deletion 34 Mb is : Supernumerary A

The value of Brain anomalies!Corpus callosum agenesis for the value of size of deletion 28 Mb is : -

The value of Brain anomalies!Aprosencephaly for the value of size of deletion 70 Mb is : -

The value of Ocular anomalies!Blepharophimosis for the value of size of deletion 47 Mb is : -

The value of Digestive anomalies!Pancreas anomalies for the value of size of deletion 70 Mb is : -

The value of Mental retardation for the value of size of deletion 34 Mb is : NK

The value of Limb defects!Thumb agenesis for the value of size of deletion 47 Mb is : +

The value of Facial dysmorphy!Hypertelorism for the value of size of deletion 47 Mb is : +

The value of Limb defects!Camptodactyly for the value of size of deletion 30 Mb is : -

The value of Brain anomalies!Cortical dysplasia for the value of size of deletion 28 Mb is : -

The value of Facial dysmorphy!Flat and broad nasal bridge for the value of size of deletion 34 Mb is : +

(1) The value of Digestive anomalies!Gall bladder agenesis for the value of size of deletion 34 Mb is : -

The value of Growth retardation for the value of size of deletion 30 Mb is : NK

The value of Facial dysmorphy!Flat and broad nasal bridge for the value of size of deletion 28 Mb is : NK

The value of Microcephaly for the value of size of deletion 47 Mb is : -

The value of Facial dysmorphy!Hypertelorism for the value of size of deletion 34 Mb is : NK

The value of Facial dysmorphy!Thick and short neck for the value of size of deletion 70 Mb is : +

The value of Displaced anus/anal atresia for the value of size of deletion 28 Mb is : +

The value of IUGR for the value of size of deletion 30 Mb is : NK

The value of Digestive anomalies!Esophageal atresia for the value of size of deletion 30 Mb is : -

The value of Ocular anomalies!Cataract for the value of size of deletion 70 Mb is : -

The value of Facial dysmorphy!Microretrognathism for the value of size of deletion 70 Mb is : NK

The value of Heart defects!Tetralogy of Fallot for the value of size of deletion 70 Mb is : -

The value of Heart defects!Tetralogy of Fallot for the value of size of deletion 34 Mb is : +

The value of Ocular anomalies!Micro/Anophthalmia for the value of size of deletion 34 Mb is : +

The value of Limb defects!Single transverse palmar crease for the value of size of deletion 30 Mb is : -

The value of Heart defects!DSCV for the value of size of deletion 34 Mb is : -

The value of Heart defects!Interauricular communication for the value of size of deletion 28 Mb is : -

The value of Brain anomalies!Holoprosencephaly for the value of size of deletion 34 Mb is : +

The value of Displaced anus/anal atresia for the value of size of deletion 70 Mb is : +

The value of Mental retardation for the value of size of deletion 47 Mb is : NK

The value of Displaced anus/anal atresia for the value of size of deletion 30 Mb is : NK

The value of Facial dysmorphy!Thick and short neck for the value of size of deletion 30 Mb is : NK

The value of Digestive anomalies!Esophageal atresia for the value of size of deletion 47 Mb is : -

The value of IUGR for the value of size of deletion 30 Mb is : +

The value of Limb defects!Camptodactyly for the value of size of deletion 34 Mb is : +

The value of Ocular anomalies!Retinal dysplasia for the value of size of deletion 47 Mb is : -

The value of Heart defects!Tetralogy of Fallot for the value of size of deletion 30 Mb is : -

The value of Facial dysmorphy!Hypotelorism for the value of size of deletion 30 Mb

is : NK

The value of Facial dysmorphy!Microretrognathism for the value of size of deletion 34 Mb is : NK

The value of Ocular anomalies!Blepharophimosis for the value of size of deletion 28 Mb is : -

The value of Limb defects!Single transverse palmar crease for the value of size of deletion 30 Mb is : -

The value of Brain anomalies!Aprosencephaly for the value of size of deletion 34 Mb is : -

The value of Limb defects!Thumb agenesis for the value of size of deletion 30 Mb is : +

The value of Limb defects!Club foot for the value of size of deletion 47 Mb is : +

The value of Brain anomalies!Cerebellar vermis hypoplasia for the value of size of deletion 47 Mb is : +

The value of Digestive anomalies!Pancreas anomalies for the value of size of deletion 34 Mb is : -

The value of Cleft lip and palate for the value of size of deletion 30 Mb is : -

The value of Limb defects!Metacarpal synostosis/ syndactyly for the value of size of deletion 70 Mb is : +

The value of Heart defects!DSCV for the value of size of deletion 47 Mb is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia for the value of size of deletion 28 Mb is : +

The value of Limb defects!Clinodactyly V for the value of size of deletion 34 Mb is : -

The value of Growth retardation for the value of size of deletion 47 Mb is : NK

The value of Brain anomalies!Aprosencephaly for the value of size of deletion 30 Mb is : -

The value of Ocular anomalies!Coloboma for the value of size of deletion 30 Mb is : -

The value of Brain anomalies!Cortical dysplasia for the value of size of deletion 70 Mb is : -

The value of Displaced anus/anal atresia for the value of size of deletion 34 Mb is : +

The value of Limb defects!Single transverse palmar crease for the value of size of deletion 70 Mb is : Unilateral A The value of Facial dysmorphy!Hypertelorism for the value of size of deletion 70 Mb is : +

The value of Brain anomalies!Corpus callosum agenesis for the value of size of deletion 34 Mb is : +

The value of Ocular anomalies!Retinal dysplasia for the value of size of deletion 70 Mb is : -

The value of Limb defects!Single transverse palmar crease for the value of size of deletion 47 Mb is : -

The value of Digestive anomalies!Spleen for the value of size of deletion 30 Mb is : NK

The value of Brain anomalies!Holoprosencephaly for the value of size of deletion 47 Mb is : -

The value of Facial dysmorphy!Microretrognathism for the value of size of deletion 47 Mb is : +

The value of Limb defects!Thumb agenesis for the value of size of deletion 70 Mb is : -

The value of Brain anomalies!Holoprosencephaly for the value of size of deletion 30 Mb is : -

The value of Brain anomalies!Holoprosencephaly for the value of size of deletion 30 Mb is : +

The value of Heart defects!Aorta coarctation for the value of size of deletion 47 Mb is : -

The value of IUGR for the value of size of deletion 70 Mb is : +

The value of Brain anomalies!Cerebellar vermis hypoplasia for the value of size of deletion 70 Mb is : +

The value of Brain anomalies!Cortical dysplasia for the value of size of deletion 34 Mb is : -

The value of Ocular anomalies!Cataract for the value of size of deletion 34 Mb is : -

The value of Facial dysmorphy!Short philtrum for the value of size of deletion 28 Mb is : NK

The value of Brain anomalies!Cortical dysplasia for the value of size of deletion 30 Mb is : NK

The value of Digestive anomalies!Spleen for the value of size of deletion 70 Mb is : -

The value of Facial dysmorphy!Malformed ears for the value of size of deletion 70 Mb is : +

The value of Facial dysmorphy!Hypotelorism for the value of size of deletion 47 Mb is : NK

The value of Ocular anomalies!Micro/Anophthalmia for the value of size of deletion 47 Mb is : -

The value of Digestive anomalies!Gall bladder agenesis for the value of size of deletion 30 Mb is : -

The value of Facial dysmorphy!Short philtrum for the value of size of deletion 34 Mb is : +

The value of Ocular anomalies!Blepharophimosis for the value of size of deletion 70 Mb is : -

The value of Microcephaly for the value of size of deletion 28 Mb is : -

The value of Facial dysmorphy!Malformed ears for the value of size of deletion 28 Mb is : +

The value of Facial dysmorphy!Flat and broad nasal bridge for the value of size of deletion 70 Mb is : NK

The value of Brain anomalies!Aprosencephaly for the value of size of deletion 30 Mb

is : -

The value of Microcephaly for the value of size of deletion 34 Mb is : +

The value of Mental retardation for the value of size of deletion 30 Mb is : NK

The value of Facial dysmorphy!Microretrognathism for the value of size of deletion 30 Mb is : NK

The value of Brain anomalies!Sylvius aqueduct dysplasia for the value of size of deletion 70 Mb is : -

The value of Digestive anomalies!Gall bladder agenesis for the value of size of deletion 47 Mb is : -

The value of Growth retardation for the value of size of deletion 28 Mb is : NK

The value of Brain anomalies!Aprosencephaly for the value of size of deletion 47 Mb is : -

The value of Genital anomalies for the value of size of deletion 28 Mb is : -

The value of Vertebral anomalies for the value of size of deletion 28 Mb is : -

The value of Facial dysmorphy!Malformed ears for the value of size of deletion 34 Mb is : +

The value of Heart defects!Aorta coarctation for the value of size of deletion 70 Mb is : -

The value of Digestive anomalies!Spleen for the value of size of deletion 47 Mb is : Hypoplasia A The value of Ocular anomalies!Coloboma for the value of size of deletion 47 Mb is : -

The value of Cleft lip and palate for the value of size of deletion 47 Mb is : +

The value of Facial dysmorphy!Hypotelorism for the value of size of deletion 30 Mb is : +

The value of Facial dysmorphy!Flat and broad nasal bridge for the value of size of deletion 47 Mb is : +

The value of Facial dysmorphy!Hypertelorism for the value of size of deletion 30 Mb is : NK

The value of Limb defects!Club foot for the value of size of deletion 30 Mb is : -

The value of Ocular anomalies!Retinal dysplasia for the value of size of deletion 28 Mb is : -

The value of Ocular anomalies!Cataract for the value of size of deletion 30 Mb is : +

The value of Facial dysmorphy!Microretrognathism for the value of size of deletion 28 Mb is : +

The value of Facial dysmorphy!Microretrognathism for the value of size of deletion 30 Mb is : NK

The value of Growth retardation for the value of size of deletion 30 Mb is : +

The value of Heart defects!ARSA for the value of size of deletion 34 Mb is : +

The value of Vertebral anomalies for the value of size of deletion 34 Mb is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia for the value of size of deletion 34 Mb is : -

The value of Heart defects!Aorta coarctation for the value of size of deletion 28 Mb is : -

The value of Ocular anomalies!Blepharophimosis for the value of size of deletion 34 Mb is : -

The value of Mental retardation for the value of size of deletion 70 Mb is : NK

The value of Facial dysmorphy!Thick and short neck for the value of size of deletion 28 Mb is : NK

The value of Heart defects!ARSA for the value of size of deletion 47 Mb is : -

The value of Limb defects!Metacarpal synostosis/ syndactyly for the value of size of deletion 34 Mb is : -

The value of Ocular anomalies!Cataract for the value of size of deletion 30 Mb is : -

The value of Limb defects!Metacarpal synostosis/ syndactyly for the value of size of deletion 28 Mb is : -

The value of Genital anomalies for the value of size of deletion 70 Mb is : +

The value of Ocular anomalies!Micro/Anophthalmia for the value of size of deletion

30 Mb is : -

The value of Genital anomalies for the value of size of deletion 34 Mb is : +

The value of Facial dysmorphy!Short philtrum for the value of size of deletion 30 Mb is : +

The value of Ocular anomalies!Blepharophimosis for the value of size of deletion 30 Mb is : -

The value of Facial dysmorphy!Flat and broad nasal bridge for the value of size of deletion 30 Mb is : NK

The value of Ocular anomalies!Coloboma for the value of size of deletion 70 Mb is : -

The value of Limb defects!Camptodactyly for the value of size of deletion 47 Mb is : -

The value of Heart defects!ARSA for the value of size of deletion 30 Mb is : -

The value of Heart defects!Aorta coarctation for the value of size of deletion 34 Mb is : -

The value of Heart defects!Interauricular communication for the value of size of deletion 70 Mb is : NK

The value of Brain anomalies!Corpus callosum agenesis for the value of size of deletion 30 Mb is : -

The value of Limb defects!Club foot for the value of size of deletion 28 Mb is : +

The value of Limb defects!Camptodactyly for the value of size of deletion 30 Mb is : -

The value of Vertebral anomalies for the value of size of deletion 47 Mb is : +

The value of Heart defects!DSCV for the value of size of deletion 70 Mb is : -

The value of Heart defects!Tetralogy of Fallot for the value of size of deletion 30 Mb is : +

The value of Mental retardation for the value of size of deletion 30 Mb is : +

The value of Limb defects!Clinodactyly V for the value of size of deletion 30 Mb is : -

The value of Digestive anomalies!Pancreas anomalies for the value of size of deletion 28 Mb is : +

The value of Heart defects!Aorta coarctation for the value of size of deletion 30 Mb is : -

The value of Ocular anomalies!Coloboma for the value of size of deletion 28 Mb is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia for the value of size of deletion 30 Mb is : NK

The value of Facial dysmorphy!Flat and broad nasal bridge for the value of size of deletion 30 Mb is : +

The value of Facial dysmorphy!Thick and short neck for the value of size of deletion 30 Mb is : NK

The value of Digestive anomalies!Pancreas anomalies for the value of size of deletion 30 Mb is : -

The value of Brain anomalies!Cortical dysplasia for the value of size of deletion 30 Mb is : -

The value of Mental retardation for the value of size of deletion 28 Mb is : NK

The value of Facial dysmorphy!Hypertelorism for the value of size of deletion 28 Mb is : +

The value of Heart defects!Interauricular communication for the value of size of deletion 34 Mb is : -

The value of Heart defects!Tetralogy of Fallot for the value of size of deletion 47 Mb is : -

The value of Limb defects!Thumb agenesis for the value of size of deletion 28 Mb is : -

The value of Limb defects!Clinodactyly V for the value of size of deletion 47 Mb is : -

The value of Heart defects!ARSA for the value of size of deletion 28 Mb is : -

The value of Heart defects!Interauricular communication for the value of size of dele-

tion 30 Mb is : -

The value of Brain anomalies!Cortical dysplasia for the value of size of deletion 47 Mb is : -

The value of Cleft lip and palate for the value of size of deletion 28 Mb is : -

The value of Ocular anomalies!Cataract for the value of size of deletion 47 Mb is : -

The value of Brain anomalies!Cerebellar vermis hypoplasia for the value of size of deletion 28 Mb is : -

The value of Vertebral anomalies for the value of size of deletion 70 Mb is : +

The value of Digestive anomalies!Pancreas anomalies for the value of size of deletion 30 Mb is : -

The value of Facial dysmorphy!Short philtrum for the value of size of deletion 70 Mb is : NK

The value of Ocular anomalies!Micro/Anophthalmia for the value of size of deletion 70 Mb is : -

The value of Facial dysmorphy!Short philtrum for the value of size of deletion 47 Mb is : +

The value of Limb defects!Clinodactyly V for the value of size of deletion 28 Mb is : +

The value of Ocular anomalies!Coloboma for the value of size of deletion 30 Mb is : -

The value of Digestive anomalies!Esophageal atresia for the value of size of deletion 70 Mb is : -

The value of Cleft lip and palate for the value of size of deletion 70 Mb is : -

The value of Digestive anomalies!Esophageal atresia for the value of size of deletion 34 Mb is : -

The value of Facial dysmorphy!Thick and short neck for the value of size of deletion 47 Mb is : +

The value of Displaced anus/anal atresia for the value of size of deletion 30 Mb is : -

The value of Vertebral anomalies for the value of size of deletion 30 Mb is : NK

The value of Digestive anomalies!Esophageal atresia for the value of size of deletion 28 Mb is : -

The value of Brain anomalies!Cerebellar vermis hypoplasia for the value of size of deletion 34 Mb is : -

The value of Heart defects!DSCV for the value of size of deletion 28 Mb is : -

The value of Digestive anomalies!Spleen for the value of size of deletion 28 Mb is : NK

The value of Brain anomalies!Cerebellar vermis hypoplasia for the value of size of deletion 30 Mb is : NK

The value of Genital anomalies for the value of size of deletion 30 Mb is : -

The value of Limb defects!Single transverse palmar crease for the value of size of deletion 28 Mb is : -

The value of Ocular anomalies!Blepharophimosis for the value of size of deletion 30 Mb is : -

The value of IUGR for the value of size of deletion 28 Mb is : -

The value of Limb defects!Thumb agenesis for the value of size of deletion 30 Mb is : -

The value of Digestive anomalies!Pancreas anomalies for the value of size of deletion 47 Mb is : -

The value of IUGR for the value of size of deletion 34 Mb is : +

The value of Ocular anomalies!Retinal dysplasia for the value of size of deletion 30 Mb is : -

The value of Heart defects!Interauricular communication for the value of size of deletion 30 Mb is : -

The value of Ocular anomalies!Retinal dysplasia for the value of size of deletion 34 Mb is : +

The value of Digestive anomalies!Spleen for the value of size of deletion 30 Mb is : -

The value of Microcephaly for the value of size of deletion 30 Mb is : -

The value of Limb defects!Clinodactyly V for the value of size of deletion 30 Mb is : -

The value of Displaced anus/anal atresia for the value of size of deletion 47 Mb is : +

The value of Vertebral anomalies for the value of size of deletion 30 Mb is : -

The value of Heart defects!ARSA for the value of size of deletion 70 Mb is : -

The value of Cleft lip and palate for the value of size of deletion 34 Mb is : +

The value of Ocular anomalies!Cataract for the value of size of deletion 28 Mb is : -

The value of Limb defects!Club foot for the value of size of deletion 34 Mb is : +

The value of Microcephaly for the value of size of deletion 30 Mb is : +

The value of Limb defects!Metacarpal synostosis/ syndactyly for the value of size of deletion 30 Mb is : -

The value of Limb defects!Metacarpal synostosis/ syndactyly for the value of size of deletion 30 Mb is : +

The value of Heart defects!DSCV for the value of size of deletion 30 Mb is : -

The value of Heart defects!Interauricular communication for the value of size of deletion 47 Mb is : -

The value of Ocular anomalies!Micro/Anophthalmia for the value of size of deletion 28 Mb is : -

The value of Digestive anomalies!Gall bladder agenesis for the value of size of deletion 30 Mb is : +

The value of Facial dysmorphy!Short philtrum for the value of size of deletion 30 Mb is : NK

The value of Facial dysmorphy!Malformed ears for the value of size of deletion 30 Mb is : NK

The value of Brain anomalies!Holoprosencephaly for the value of size of deletion 28 Mb is : -

The value of Brain anomalies!Sylvius aqueduct dysplasia for the value of size of dele-

tion 30 Mb is : -

The value of Digestive anomalies!Gall bladder agenesis for the value of size of deletion 28 Mb is : -

The value of Facial dysmorphy!Malformed ears for the value of size of deletion 30 Mb is : +

The value of IUGR for the value of size of deletion 47 Mb is : +

The value of Limb defects!Camptodactyly for the value of size of deletion 28 Mb is : -

The value of Facial dysmorphy!Hypotelorism for the value of size of deletion 70 Mb is : NK

The value of Brain anomalies!Corpus callosum agenesis for the value of size of deletion 47 Mb is : +

The value of Brain anomalies!Holoprosencephaly for the value of size of deletion 70 Mb is : -

The value of Growth retardation for the value of size of deletion 70 Mb is : NK

The value of Heart defects!Tetralogy of Fallot for the value of size of deletion 28 Mb is : -

The value of Limb defects!Single transverse palmar crease for the value of size of deletion 34 Mb is : -

The value of Brain anomalies!Corpus callosum agenesis for the value of size of deletion 30 Mb is : NK

The value of Facial dysmorphy!Hypotelorism for the value of size of deletion 28 Mb is : NK

The value of Limb defects!Thumb agenesis for the value of size of deletion 34 Mb is : -

The value of Microcephaly for the value of size of deletion 70 Mb is : +

The value of Ocular anomalies!Coloboma for the value of size of deletion 34 Mb is : +

The value of Limb defects!Club foot for the value of size of deletion 70 Mb is : -

The value of Facial dysmorphy!Hypertelorism for the value of size of deletion 30 Mb is : +

The value of Limb defects!Camptodactyly for the value of size of deletion 70 Mb is : -

The value of Digestive anomalies!Gall bladder agenesis for the value of size of deletion 70 Mb is : -

The value of Ocular anomalies!Micro/Anophthalmia for the value of size of deletion 30 Mb is : -

The value of Ocular anomalies!Retinal dysplasia for the value of size of deletion 30 Mb is : -

The value of Facial dysmorphy!Malformed ears for the value of size of deletion 47 Mb is : +

The value of Genital anomalies for the value of size of deletion 30 Mb is : +

The value of Growth retardation for the value of size of deletion 34 Mb is : NK

The value of Brain anomalies!Aprosencephaly for the value of size of deletion 28 Mb is : -

The value of Heart defects!ARSA for the value of size of deletion 30 Mb is : -

The value of Limb defects!Club foot for the value of size of deletion 30 Mb is : -

The value of Limb defects!Metacarpal synostosis/ syndactyly for the value of size of deletion 47 Mb is : -

The value of Brain anomalies!Cerebellar vermis hypoplasia for the value of size of deletion 30 Mb is : -

The value of Heart defects!DSCV for the value of size of deletion 30 Mb is : +

The value of Genital anomalies for the value of size of deletion 47 Mb is : +

The value of Brain anomalies!Sylvius aqueduct dysplasia for the value of size of deletion 47 Mb is : -

The value of Facial dysmorphy!Hypotelorism for the value of size of deletion 34 Mb

is : +

The value of Heart defects!Aorta coarctation for the value of size of deletion 30 Mb is : -

The value of Limb defects!Clinodactyly V for the value of size of deletion 70 Mb is : +

The value of Brain anomalies!Corpus callosum agenesis for the value of size of deletion 70 Mb is : -

The value of Facial dysmorphy!Thick and short neck for the value of size of deletion 34 Mb is : NK

The value of Digestive anomalies!Esophageal atresia for the value of size of deletion 30 Mb is : -

Appendix B

Result of Query VIIB

B.1 Result of Query VIIB

For the clinical phenotype NC/SV the corresponding mutation value and gene expression is p.R233K and $g.1586G > A$ respectively.

For the clinical phenotype SW the corresponding mutation value and gene expression is 9bpins exon 2 and $g.519\_520insTGTGGTGGT$ respectively.

For the clinical phenotype Normal the corresponding mutation value and gene expression is p.A265V and $g.1851C > T$ respectively.

For the clinical phenotype SV the corresponding mutation value and gene expression is p.E320K and $g.2216G > A$ respectively.

For the clinical phenotype NC/SV the corresponding mutation value and gene expression is p.K121Q and $g.952A > C$ respectively.

For the clinical phenotype SW the corresponding mutation value and gene expression is p.T450P and $g.2786A > C$ respectively.

For the clinical phenotype SW the corresponding mutation value and gene expression is p.L142P and $g.1016T > C$ respectively.

For the clinical phenotype SV the corresponding mutation value and gene expression

is p.W302S and $g.1962G > C$ respectively.

For the clinical phenotype NC the corresponding mutation value and gene expression is p.D407N and $g.2560G > A$ respectively.

For the clinical phenotype Normalb the corresponding mutation value and gene expression is p.V249A and $g.1803T > C$ respectively.

For the clinical phenotype NC the corresponding mutation value and gene expression is p.N387K and $g.2502C > G$ respectively.

For the clinical phenotype NC the corresponding mutation value and gene expression is p.I230T and $g.1577T > C$ respectively.

For the clinical phenotype SW the corresponding mutation value and gene expression is p.L353R and $g.2316T > G$ respectively.

For the clinical phenotype SW the corresponding mutation value and gene expression is p.L107R and $g.911T > G$ respectively.

For the clinical phenotype SV the corresponding mutation value and gene expression is p.G56R and $g.378G > A$ respectively.

For the clinical phenotype CL the corresponding mutation value and gene expression is p.Y59N and $g.387T > A$ respectively.

For the clinical phenotype SV the corresponding mutation value and gene expression is 10bpdel exon 1 and $g.231\_240delCTGCTGCTGC$ respectively.

For the clinical phenotype SW the corresponding mutation value and gene expression is p.W22X and $g.277G > A$ respectively.

For the clinical phenotype SW the corresponding mutation value and gene expression is p.K54X and $g.372A > T$ respectively.

For the clinical phenotype CL the corresponding mutation value and gene expression is p.M1V and $g.213A > G$ respectively.

For the clinical phenotype NC the corresponding mutation value and gene expression is p.D322G and $g.2223A > G$ respectively.

For the clinical phenotype SW the corresponding mutation value and gene expression

is p.F404S and $g.2552T > C$ respectively.

For the clinical phenotype SW the corresponding mutation value and gene expression is p.G292D and $g.1931G > A$ respectively.

For the clinical phenotype SW the corresponding mutation value and gene expression is p.L167P and $g.1199T > C$ respectively.

For the clinical phenotype NC the corresponding mutation value and gene expression is p.R224W and $g.1558C > T$ respectively.

For the clinical phenotype NC/SV the corresponding mutation value and gene expression is p.R369W and $g.2363C > T$ respectively.

For the clinical phenotype NC the corresponding mutation value and gene expression is p.I194N and $g.1368T > A$ respectively.

For the clinical phenotype NC the corresponding mutation value and gene expression is p.H119R and $g.947A > G$ respectively.

Appendix C

Result of Query of Table 1

C.1 Values of the score

13
10
6
8
0
5
19
16
15
12
7
18
9
14
21

6

17

11

C.2 Values of the cDNA

c.2381dupC

Duplication

c.625C4T

c.1828delG

c.925-1G4A

c.1123-1G4A

Large deletion

c.2862C4T

Nonsense

Insertion

RNA processing

$c.1311_1313delinsCA$

c.1693C4T

c.3228delT

c.2518-1 G4A

c.3137dupC

c.3112-1G4A

c.3106C4T

c.744delT

c.1777C4T

$c.211_233dup$

c.4074+1 G4T

c.2193+2T4C

c.3003+1G4A

c.2813delC

c.2673dupC

c.1888-2A4G

c.1172delC

c.3574C4T

Deletion/insertion

c.1957C4T

c.2094+1G4A

c.2710C4T

c.2715dupT

c.1931delC

$c.264_276del$

c.2588-2604delCTGG TCCTCAGGGCCCC

c.654+1G4A

c.655G4C

c.430-1G4C

c.3003+5G4A

c.2355+5G4A

del COL2A1

c.3325delC

c.1833+1G4A

c.870+5G4A

c.665G4T

c.3641dupC

Deletion

c.3138delT

c.2095-1G4A

Synonymous

c.647G4A

c.2493dupA

c.2467G4T

c.3392G4C

c.2673delC

c.2101C4T

c.3137delC

c.3864-3865delCT

$c.3258_3261delAGAC$

c2355+5G4A

c.3623delC

c.2353C4T

Missense

$c.3891_3898dupCTACTGGA$

c.1221+1G4A

c.2382delT

c.1680+2delGTinsAA

c.1030C4T

c.793delG

c.1833+1 G4A

c.2659C4T

c.1475G4A

RNA processing a

$c.3081_3087delGACGGTinsCCTGG$

c.4317+2T4C

c.3903delC

$c.1428_1429insTGGC$

c.3111+1G4T

c.1597C4T

$c.2257_2264delGGCGAGAG$

c.2719dupC

c.492delT

$c.2263_2264delAG$

c.3878G4A

c.2839C4T

c.2517+2T4G

c.342+1G4A

C.3 Values of the Mutation Type

Nonsense

Large deletion

Deletion

Duplication

RNA processing a

Deletion/insertion

Insertion

RNA processing

Missense

Synonymous

C.4 Values of the Mutation Effect

Insertion with premature stop codon

Deletion[19]

Glycine substitution

Arginine-to-cysteine substitution[20]

Insertion with premature stop codon

ND

Frameshift

Deletion frameshift

Frameshift, insertion with premature stop codon,exon deletion

frameshift

Premature stop codon

C.5 Values of the Protein

p.Glu823X

p.Gly492Asp

p.Ala895SerfsX49

p.Gly1077AlafsX53

p.Glu754SerfsX13

p.Thr1028LeufsX100

p.Gly795Alafs86

$p.Gly144ValfsX54p.Gln125_{G}ly126insArgGluGlyGluAsnLeuPheLeuArgPro$

$aPheLeuAlaAlaGlnValThrAspLeuX20p.Lys143_{A}sn178delExon7a$

p.Gly906TrpfsX38

p.Gln1109ArgfsX21

p.Gly954Glyb

ND

$p.Gln1238_{L}eu1411delp.Trp1348CysfsX17a$

p.Arg209X

p.Cys1289ProfsX3

p.Ile1300ThrfsX15

del COL2A1

p.Glu1033LysfsX4

p.Arg755GlyfsX14

p.Gly165ValfsX34

p.Gly375ValfsX253

p.Pro893ArgfsX135

p.Pro832ThrfsX11

p.Gly222Val

p.Gly438ThrfsX191

p.Arg565Cys

$p.Asp114_{I}le115insIleSerAlaAsnTyrSerHisProValLeuGlnLeuLeuX14$

p.Gly795TrpfsX6

p.Gly1215TrpfsX38

p.Ala610ProfsX19

p.Gly630MetfsX53

p.Pro1046LeufsX84

p.Pro938LeufsX90

p.Gln947X

p.Arg701X

p.Gly609GlyfsX1

p.Pro391LeufsX238

p.Gly1047AlafsX83

p.Gln593X

p.Glu79ThrfsX2

p.Gly216Asp

p.Arg1192X

p.Cys89SerfsX24

p.Arg904Cys

p.Arg887X

$p.Lys308_Gly309insGluPheAlaGlyGlyGlnGluTrpGlyProArgHisX13$

p.Gly1131Ala

p.Gly1047TrpfsX11

p.Gly477TrpfsX12

p.Pro863LeufsX16

p.Gly249GlufsX59

p.Asn1303ThrfsX9

p.Arg653X

p.Asp1087GlufsX42

p.Arg1036X

p.Arg785X

p.Gly219Arg

p.Trp1293X

p.Arg344X

p.Pro1208LeufsX19

p.Gly909ArgfsX35

$p.Arg785_Gly786insValAsnAlaPheGlySerX15$

p.Pro644LeufsX144

p.Gly840ValfsX41

p.Arg533X

p.Gly1038GlufsX92

p.Glu265fsX43

$p.Gly966_ser1001delp.Gly990Glyfs X1a$

C.6 Values of the Patient Id

50

36

81

9

16

94

29

33

45

84

77

63

97

2

52

87

83

46

18

22

24

66

53
79
6
31
41
43
48
13
19
26
38
20
60
91
1
30
59
89
39
72
15
12
25
96
67
69
80
93

32
5
58
7
76
34
14
85
78
98
82
35
21
47
65
10
99
44
90
54
51
86
49
4
55
71
73
11

28
8
40
42
23
68
57
62
37
75
17
92
88
27
56
3
64
61
100
95
74
70

C.7 Values of the Exon/Intron

38
51

IVS 43
35
EMPTY
19
40
48
IVS 32
17
25
IVS 13
09
29
IVS 35
21
IVS 06
50
IVS 28
i
10
IVS 18
IVS 14
47
07
IVS 25
02
46
42
IVS 38

34

39

IVS 44

09

IVS 19

IVS 33

33

11

IVS 09

IVS 53

45

26

IVS 52

30

44

27

23

IVS 27

12

52

36

IVS 04

41

C.8 Values of the Age(Years)

19
58
6
45
47
27
40
36
13
30
24
43
6
8
34
11
38
32
55
70
10
3
4
44
67
54
46

18

9

29

35

39

31

22

49

42

5

41

52

12

66

33

17

18 (8 at examination)

37

20

62

14

Appendix D

Result of Query of Table 2

D.1 Nucleotide Change

1152-1189del

1144-1193del

905 C ! T

880 C ! T

916 C ! T

808 C ! T

473 C ! T

863-881del

1158-1201del

76delC

1163-1188del

502 C ! T

763 C ! T

423 C ! G

397 C ! T

1169-1197del

316 C ! T

1164-1217del

766dup14

1156-1199del

D.2 Phenotypes

At

nk

At, PSV

4 Cl

5 Cl

Cl

3 Cl

2 Cl

5 Cl, 2 At

2 Cl, 1 nk

5 Cl, 1 At

D.3 Type of Mutation

R133C

Y141X

P302L

R106W

R168X

R306C

T158M

Frameshift

R255X

R294X

R270X

D.4 Number of Patients

1

2

7

4

5

3

6

Appendix E

Reference Papers

E.1 Reference Papers

TABLE 2. Genotype-phenotype characteristics of 9 probands with Kallmann syndrome due to KAL mutations from [The Importance of Autosomal Genes in Kallmann Syndrome: Genotype-Phenotype Correlations and Neuroendocrine Characteristics Authors Luciana M. B. Oliveira, Stephanie B. Seminara, Milena Beranova, Frances J. Hayes, Sarah B. Valkenburgh, Ernestina Schipani, Elaine Maria F. Costa, Ana Claudia Latronico, William F. Crowley, Jr., and Mario Vallejo] Table 1. Summary of Clinical Findings from [Deletions at the SOX10 Gene Locus Cause Waardenburg Syndrome Types 2 and 4 Author(s) Nadege Bondurand, Florence Dastot-Le Moal, Laure Stanchina, Nathalie Collot, Viviane Baral, Sandrine Marlin, Tania Attie-Bitach, Irina Giurgea, Laurent Skopinski, William Reardon, Annick Toutain, Pierre Sarda, Anis Echaieb, Marilyn Lackmy-Port-Lis, Renaud Touraine, Jeanne Amiel, Michel Goossens, and Veronique Pingault]

Table 2. QMF-PCR Primer Sequences from [Deletions at the SOX10 Gene Locus Cause Waardenburg Syndrome Types 2 and 4 Author(s) Nadege Bondurand, Florence

Dastot-Le Moal, Laure Stanchina, Nathalie Collot, Viviane Baral, Sandrine Marlin, Tania Attie-Bitach, Irina Giurgea, Laurent Skopinski, William Reardon, Annick Toutain, Pierre Sarda, Anis Echaieb, Marilyn Lackmy-Port-Lis, Renaud Touraine, Jeanne Amiel, Michel Goossens, and Veronique Pingault]

Table 4. Additional QMF-PCR Primers, Ordered Centromeric to Telomeric from [Deletions at the SOX10 Gene Locus Cause Waardenburg Syndrome Types 2 and 4 Author(s) Nadege Bondurand, Florence Dastot-Le Moal, Laure Stanchina, Nathalie Collot, Viviane Baral, Sandrine Marlin, Tania Attie-Bitach, Irina Giurgea, Laurent Skopinski, William Reardon, Annick Toutain, Pierre Sarda, Anis Echaieb, Marilyn Lackmy-Port-Lis, Renaud Touraine, Jeanne Amiel, Michel Goossens, and Veronique Pingault]

TABLE I. Clinical and Radiographic Data of the Patients from [Novel and Recurrent EBP Mutations in X-Linked Dominant Chondrodysplasia Punctata Author(s) Shiro Ikegawa, Hirofumi Ohashi, Tsutomu Ogata, Akira Honda, Masato Tsukahara, Toshihide Kubo, Mamori Kimizuka, Masanori Shimode, Tomonobu Hasegawa, Gen Nishimura, and Yusuke Nakamura]

TABLE II. EBP Mutations in Chondrodysplasia Punctata from [Novel and Recurrent EBP Mutations in X-Linked Dominant Chondrodysplasia Punctata Author(s) Shiro Ikegawa, Hirofumi Ohashi, Tsutomu Ogata, Akira Honda, Masato Tsukahara, Toshihide Kubo, Mamori Kimizuka, Masanori Shimode, Tomonobu Hasegawa, Gen Nishimura, and Yusuke Nakamura]

Table: Phenotypic-Genotypic Comparison Between Our Case and Two Previously Reported Cases from [FGFR3 Gene Mutation (Gly380Arg) With Achondroplasia and i(21q) Down Syndrome: Phenotype-Genotype Correlation Author(s) HAROLD CHEN; JOSE MARTINEZ, CATHY TUCK-MULLER, and WLADIMIR WERTELECKI]

TABLE 1. Primers Used for Quantitative Polymerase Chain Reaction from [Identification of Candidate Genes for Congenital Ventricular Septal Defects With HSA22q11

Loss of Heterozygosity Author(s) Cheng-Liang Lee, Kai-Sheng Hsieh, Yi-Ling Chen, and Yow-Ling Shiueb]

TABLE 2. Number and Percentage of Patients With HSA22q11 Loss of Heterozygosity (LOH) by Microsatellite Genotyping in Patients With Congenital Heart Diseases from [Identification of Candidate Genes for Congenital Ventricular Septal Defects With HSA22q11 Loss of Heterozygosity Author(s) Cheng-Liang Lee, Kai-Sheng Hsieh Yi-Ling Chen and Yow-Ling Shiueb]

Table 1. Scoring System for Evaluation of Clinical Features from [APOE ε4: A Potential Modulation Factor in Rett Syndrome Daniela Zahorakova, Marie Jachymova, David Kemlink, Alice Baxova, and Pavel Martasek]

Table 2. Comparison of the Clinical Phenotype in ε4 Carriers and Noncarriers With Rett Syndrome from [APOE ε4: A Potential Modulation Factor in Rett Syndrome Daniela Zahorakova, Marie Jachymova, David Kemlink, Alice Baxova, and Pavel Martasek]

Table 1. Two-Point LOD Scores for 16 X-Chromosomal Markers from [OFD1 Is Mutated in X-Linked Joubert Syndrome and Interacts with LCA5-Encoded Lebercilin Karlien L.M. Coene, Ronald Roepman, Dan Doherty, Bushra Afroze, Hester Y. Kroes, Stef J.F. Letteboer, Lock H. Ngu,5 Bartlomiej Budny, Erwin van Wijk, Nicholas T. Gorden, Malika Azhimi, Christel Thauvin-Robinet, Joris A. Veltman, Mireille Boink, Tjitske Kleefstra, Frans P.M. Cremers, Hans van Bokhoven, and Arjan P.M. de Brouwer]

Table 2. Clinical Features of the XL-JS Family and Isolated JS Patient UW87 as Compared to SGBS2 and Females with OFD1 Syndrome from [OFD1 Is Mutated in X-Linked Joubert Syndrome and Interacts with LCA5-Encoded Lebercilin Karlien L.M. Coene, Ronald Roepman, Dan Doherty, Bushra Afroze, Hester Y. Kroes, Stef J.F. Letteboer,1 Lock H. Ngu, Bartlomiej Budny, Erwin van Wijk, Nicholas T. Gorden, Malika Azhimi,1 Christel Thauvin-Robinet, Joris A. Veltman, Mireille Boink,1 Tjitske Kleefstra, Frans P.M. Cremers, Hans van Bokhoven, and Arjan P.M.

de Brouwer]

TABLE II. Prevalence of Morbidity in RTT Persons Per Age-Group and Significance Level Aging in People With Specific Genetic Syndromes: Rett Syndrome from [Aging in People With Specific Genetic Syndromes: Rett Syndrome Author(s) Nicky S.J. Halbach, Eric E.J. Smeets, Connie T.R.M. Schrander-Stumpel, Henny H.J. van Schrojenstein Lantman de Valk, Marian A. Maaskant, and Leopold M.G. Curfs]

TABLE III. OOB-Scores in Relation to Age from [Aging in People With Specific Genetic Syndromes: Rett Syndrome Author(s) Nicky S.J. Halbach, Eric E.J. Smeets, Connie T.R.M. Schrander-Stumpel, Henny H.J. van Schrojenstein Lantman de Valk, Marian A. Maaskant,3,6,7 and Leopold M.G. Curfs]

Table 1. Genetic Types of the LQTS from [Long QT Syndrome and Short QT Syndrome Wojciech Zareba and Iwona Cygankiewicz]

TABLE I. Clinical and Laboratory Findings in Seven Patients with SimpsonGolabiBehmel Syndrome (SGBS) from [GPC3 Mutations in Seven Patients With SimpsonGolabiBehmel Syndrome Author(s) Satoru Sakazume, Nobuhiko Okamoto, Toshiyuki Yamamoto, Kenji Kurosawa, Hironao Numabe, Yuko Ohashi, Yuko Kako, Toshiro Nagai, and Hirohumi Ohashi]

TABLE II. Primers Used for PCR and Sequencing from [GPC3 Mutations in Seven Patients With SimpsonGolabiBehmel Syndrome Satoru Sakazume, Nobuhiko Okamoto, Toshiyuki Yamamoto, Kenji Kurosawa, Hironao Numabe, Yuko Ohashi, Yuko Kako, Toshiro Nagai, and Hirohumi Ohashi]

TABLE III. GPC3 Mutations Detected in Seven Patients with SGBS from [GPC3 Mutations in Seven Patients With SimpsonGolabiBehmel Syndrome Satoru Sakazume, Nobuhiko Okamoto, Toshiyuki Yamamoto, Kenji Kurosawa, Hironao Numabe, Yuko Ohashi, Yuko Kako, Toshiro Nagai, and Hirohumi Ohashi]

Table 2. Twenty-Seven ASD Cases with De Novo Rearrangements from [Structural Variation of Chromosomes in Autism Spectrum Disorder Author(s) Christian R. Marshall, Abdul Noor, John B. Vincent, Anath C. Lionel, Lars Feuk, Jennifer Skaug,

Mary Shago, Rainald Moessner, Dalila Pinto, Yan Ren, Bhooma Thiruvahindrapduram, Andreas Fiebig, Stefan Schreiber, Jan Friedman, Cees E.J. Ketelaars, Yvonne J. Vos, Can Ficicioglu, Susan Kirkpatrick, Rob Nicolson, Leon Sloman, Anne Summers, Clare A. Gibbons, Ahmad Teebi, David Chitayat, Rosanna Weksberg, Ann Thompson, Cathy Vardy, Vicki Crosbie, Sandra Luscombe, Rebecca Baatjes, Lonnie Zwaigenbaum, Wendy Roberts, Bridget Fernandez, Peter Szatmari, and Stephen W. Scherer]

Table 3. Recurrent and Overlapping Loci in ASD from [Structural Variation of Chromosomes in Autism Spectrum Disorder Author(s) Christian R. Marshall, Abdul Noor, John B. Vincent, Anath C. Lionel, Lars Feuk, Jennifer Skaug, Mary Shago, Rainald Moessner, Dalila Pinto, Yan Ren, Bhooma Thiruvahindrapduram, Andreas Fiebig, Stefan Schreiber, Jan Friedman, Cees E.J. Ketelaars, Yvonne J. Vos, Can Ficicioglu, Susan Kirkpatrick, Rob Nicolson, Leon Sloman, Anne Summers, Clare A. Gibbons, Ahmad Teebi, David Chitayat, Rosanna Weksberg, Ann Thompson, Cathy Vardy, Vicki Crosbie, Sandra Luscombe, Rebecca Baatjes, Lonnie Zwaigenbaum, Wendy Roberts, Bridget Fernandez, Peter Szatmari, and Stephen W. Scherer]

TABLE I. Phenotype and Genotype in 10 Patients With a COL11A1 Mutation from [A Report on 10 New Patients With Heterozygous Mutations in the COL11A1 Gene and a Review of GenotypePhenotype Correlations in Type XI Collagenopathies Author(s) Marja Majava,1 Kristien P. Hoornaert, Deborah Bartholdi, Mieke C. Bouma, Katelijne Bouman, Marta Carrera, Koenraad Devriendt, Jane Hurst, George Kitsos, Dunja Niedrist, Michael B. Petersen, Debbie Shears, Irene Stolte-Dijkstra, J.M. Van Hagen, Leena Ala-Kokko, Minna Mannikko, and Geert R. Mortier]

TABLE I. Anthropometrica Data for the Patient from [Clinical Report SmithMagenis Syndrome and Moyamoya Disease in a Patient With del(17)(p11.2p13.1) Author(s) Santhosh Girirajan, Roberto Mendoza-Londono, Christopher N. Vlangos, Lucie Dupuis, Norma J. Nowak, David J. Bunyan, Eli Hatchwell, and Sarah H. Elsea]

TABLE II. MLPA Probes Spanning RAI1 from [Clinical Report SmithMagenis Syn-

drome and Moyamoya Disease in a Patient With del(17)(p11.2p13.1) Author(s) Santhosh Girirajan, Roberto Mendoza-Londono, Christopher N. Vlangos, Lucie Dupuis, Norma J. Nowak, David J. Bunyan, Eli Hatchwell, and Sarah H. Elsea]

TABLE III. SmithMagenis Syndrome Manifestations from [Clinical Report Smith-Magenis Syndrome and Moyamoya Disease in a Patient With del(17)(p11.2p13.1) Author(s) Santhosh Girirajan, Roberto Mendoza-Londono, Christopher N. Vlangos, Lucie Dupuis, Norma J. Nowak, David J. Bunyan, Eli Hatchwell, and Sarah H. Elsea]

Table 2. XCI Ratio, Genotype, Phenotype, and Parental Origin of the Priority Inactive X Chromosome in 8 Cases With Extremely Skewed XCI from [X Chromosome Inactivation in Rett Syndrome and Its Correlations With MeCP2 Mutations and Phenotype Xinhua Bao, Shengling Jiang, Fuying Song, Hong Pan, Meirong Li, and Xi-Ru Wu]

Table 1. Summary of Genotype-Phenotype Studies Performed in Patients With Rett Syndrome or Other MECP2-Related Disorders from [Does Genotype Predict Phenotype in Rett Syndrome? Author(s) Andrea L. Ham; Asmita Kumar; Rose Deeter; N. Carolyn Schanen]

Table 2. Summary of Mutations Reported in Mild Rett Syndrome Variants from [Does Genotype Predict Phenotype in Rett Syndrome? Author(s) Andrea L. Ham; Asmita Kumar; Rose Deeter; N. Carolyn Schanen]

Table 1 Clinical manifestations of 5p- syndrome patients. IUGR intrauterine growth retardation, MR mental retardation, GR (postnatal) growth retardation, AG ataxic gait, UTW unable to walk, CHD congenital heart defect from [Genotype-phenotype correlation of 5p- syndrome: pitfall of diagnosis Author(s) Tatsuro Kondoh, Osamu Shimokawa Naoki Harada Tomoki Doi Chyuns Yun, Yuji Gohda Fumiko Kinoshita, Tadashi Matsumoto Hiroyuki Moriuchi]

Table 1. Clinical and Cytogenetic Finding among Six Patients with 22q11.2 Distal Deletion from [22q11.2 Distal Deletion: A Recurrent Genomic Disorder Distinct from DiGeorge Syndrome and Velocardiofacial Syndrome Shay Ben-Shachar, Zhishuo Ou,

Chad A. Shaw, John W. Belmont, Millan S. Patel, Marybeth Hummel, Stephen Amato, Nicole Tartaglia, Jonathan Berg, V. Reid Sutton, Seema R. Lalani, A. Craig Chinault, Sau W. Cheung, James R. Lupski, and Ankita Patel]

TABLE 1 Mendelian Diseases of Dental Importancefrom [GENES AND GENE POLYMORPHISMS ASSOCIATED WITH PERIODONTAL DISEASE D.F. Kinane University of Louisville School of Dentistry, Louisville, KY 40292, USA; corresponding author, denis.kinane@louisville.edu T.C. Hart Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, PA, USA]

TABLE 2 Examples of Syndromic Forms of Periodontitis in Which Inheritance is Mendelian and Due to a Genetic Alteration at a Single Gene Locus from [GENES AND GENE POLYMORPHISMS ASSOCIATED WITH PERIODONTAL DISEASE D.F. Kinane University of Louisville School of Dentistry, Louisville, KY 40292, USA; corresponding author, denis.kinane@louisville.edu T.C. Hart Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, PA, USA]

TABLE I. Major Features of the Patients and PhenotypeGenotype Correlations from [Mutational and GenotypePhenotype Correlation Analyses in 28 Polish Patients With Cornelia de Lange Syndrome Author(s) Jiong Yan, Gulam Mustafa Saifi, Tomasz H. Wierzba, Marjorie Withers, Gabriel A. Bien-Willner, Janusz Limon, Pawe Stankiewicz, James R. Lupski, and Jolanta Wierzba]

TABLE IIA. Summary of NIPBL Mutations from [Mutational and GenotypePhenotype Correlation Analyses in 28 Polish Patients With Cornelia de Lange Syndrome Author(s) Jiong Yan, Gulam Mustafa Saifi, Tomasz H. Wierzba, Marjorie Withers, Gabriel A. Bien-Willner, Janusz Limon, Pawe Stankiewicz, James R. Lupski, and Jolanta Wierzba]

TABLE IIB. Nucleotide Repeats at the Sites of Mutations from [Mutational and GenotypePhenotype Correlation Analyses in 28 Polish Patients With Cornelia de Lange Syndrome Author(s) Jiong Yan, Gulam Mustafa Saifi, Tomasz H. Wierzba, Marjorie Withers, Gabriel A. Bien-Willner, Janusz Limon, Pawe Stankiewicz, James

R. Lupski, and Jolanta Wierzba]

TABLE III. Summary of NIPBL Polymorphisms from [Mutational and Genotype-Phenotype Correlation Analyses in 28 Polish Patients With Cornelia de Lange Syndrome Author(s) Jiong Yan, Gulam Mustafa Saifi, Tomasz H. Wierzba, Marjorie Withers, Gabriel A. Bien-Willner, Janusz Limon, Pawe Stankiewicz, James R. Lupski, and Jolanta Wierzba]

TABLE I. Clinical Features of Families A and B, EEC Syndrome, LMS and RHS from from [Clinical Report EEC Syndrome, Arg227Gln TP63 Mutation and Micturition Difficulties: Is There a GenotypePhenotype Correlation? Author(s) Kenneth Maclean, Stephen A. Holme, Elizabeth Gilmour, Mark Taylor, Heide Scheffer, Nicole Graf, Grahame H.H. Smith, Ella Onikul, Hans van Bokhoven, Celia Moss, and Lesley C. Ade's]

Table 1 Karyotypes and clinical phenotypes of cohort A from [Genomic analysis of partial 21q monosomies with variable phenotypes Elisha DO Roberson, Elizabeth Squibb Wohler, Julie E Hoover-Fong, Emily Lisi, Eric L Stevens, George H Thomas, Jay Leonard, Ada Hamosh and Jonathan Pevsner]

Table 2 Karyotypes and clinical phenotypes of cohort B from [Genomic analysis of partial 21q monosomies with variable phenotypes Elisha DO Roberson, Elizabeth Squibb Wohler, Julie E Hoover-Fong, Emily Lisi, Eric L Stevens, George H Thomas, Jay Leonard, Ada Hamosh and Jonathan Pevsner]

Table 1. Clinical Features of CED Patients with and without IFT122 Mutations from [Cranioectodermal Dysplasia, Sensenbrenner Syndrome, Is a Ciliopathy Caused by Mutations in the IFT122 Gene Joanna Walczak-Sztulpa, Jonathan Eggenschwiler, Daniel Osborn, Desmond A. Brown, Francesco Emma, Claus Klingenberg, Raoul C. Hennekam, Giuliano Torre, Masoud Garshasbi, Andreas Tzschach, Malgorzata Szczepanska, Marian Krawczynski, Jacek Zachwieja, Danuta Zwolinska, Philip L. Beales, Hans-Hilger Ropers, Anna Latos-Bielenska, and Andreas W. Kuss]

TABLE 1 A-T Patients with ATM Protein from [Ataxia-Telangiectasia: Pheno-

type/Genotype Studies of ATM Protein Expression, Mutations, and Radiosensitivity Sara G. Becker-Catania, Gang Chen, Mee Jeong Hwang, Zhijun Wang, Xia Sun, Ozden Sanal, Eva Bernatowska-Matuszkiewicz, Luciana Chessa,Eva Y.-H.P. Lee, and Richard A. Gatti]

TABLE 2 Patients Referred with Suspected Diagnosis of A-T from [Ataxia-Telangiectasia: Phenotype/Genotype Studies of ATM Protein Expression, Mutations, and Radiosensitivity Sara G. Becker-Catania, Gang Chen, Mee Jeong Hwang, Zhijun Wang, Xia Sun, Ozden Sanal, Eva Bernatowska-Matuszkiewicz, Luciana Chessa, Eva Y.-H.P. Lee, and Richard A. Gatti]

Table 1 Summary of age-corrected cognitive performance (in italics poor or impaired scores) and extraocular findings from [Extended extraocular phenotype of PROM1 mutation in kindreds with known autosomal dominant macular dystrophy Francesca I Arrigoni, Mar Matarin, Pamela J Thompson, Michel Michaelides, Michelle E McClements, Elizabeth Redmond, Lindsey Clarke, Elizabeth Ellins, Saifullah Mohamed, Ian Pavord, David M Hunt, Anthony T Moore, Julian Halcox and Sanjay M Sisodiya]

Table 2 Blood/vasculature tests in controls and PROM1 subjects from [Extended extraocular phenotype of PROM1 mutation in kindreds with known autosomal dominant macular dystrophy Francesca I Arrigoni, Mar Matarin, Pamela J Thompson, Michel Michaelides, Michelle E McClements, Elizabeth Redmond, Lindsey Clarke, Elizabeth Ellins, Saifullah Mohamed, Ian Pavord, David M Hunt, Anthony T Moore, Julian Halcox, and Sanjay M Sisodiya]

Table 1 Oligonucleotide Primers for Amplification of the NAGLU Coding Region in Four Segments from [Genotype-Phenotype Correspondence in Sanfilippo Syndrome Type B Author(s) Hong G. Zhao, Elena L. Aronovich and Chester B. Whitley]

Table 2 Oligonucleotide Primers for Automated Sequencing of the NAGLU Coding Region from [Genotype-Phenotype Correspondence in Sanfilippo Syndrome Type B Author(s) Hong G. Zhao, Elena L. Aronovich and Chester B. Whitley]

Table 3 Summary of Pathological NAGLU Mutations and Clinical Phenotype in Pa-

tients with Sanfilippo Syndrome Type B from [Genotype-Phenotype Correspondence in Sanfilippo Syndrome Type B Author(s) Hong G. Zhao, Elena L. Aronovich and Chester B. Whitley]

Table 4 Mutations of the NAGLU Gene from [Genotype-Phenotype Correspondence in Sanfilippo Syndrome Type B Author(s) Hong G. Zhao, Elena L. Aronovich and Chester B. Whitley]

TABLE I. Clinical Findings in Our Patient and Those in CFC and CS from [Clinical Report Hepatoblastoma and Heart Transplantation in a Patient With Cardio-Facio-Cutaneous Syndrome Author(s) Mohamad M. Al-Rahawan, Deborah J. Chute, Katia Sol-Church, Karen W. Gripp, Deborah L. Stabley, Nancy L. McDaniel, William G. Wilson, and Peter E. Waldron]

TABLE II. PCR and Sequencing Primers Used in This Case from [Clinical Report Hepatoblastoma and Heart Transplantation in a Patient With Cardio-Facio-Cutaneous Syndrome Mohamad M. Al-Rahawan, Deborah J. Chute, Katia Sol-Church, Karen W. Gripp, Deborah L. Stabley, Nancy L. McDaniel, William G. Wilson, and Peter E. Waldron]

Table 1 Distinctive phenotype groups in FA and examples of corresponding mutations from [Genotypephenotype correlations in Fanconi anemia Kornelia Neveling, Daniela Endt, Holger Hoehn, Detlev Schindler]

Table 1 HLXB9 Mutations Identified in the Study and Associated Phenotypes from [Mutation Analysis and Embryonic Expression of the HLXB9 Currarino Syndrome Gene D. M. Hagan, A. J. Ross, T. Strachan, S. A. Lynch, V. Ruiz-Perez, Y. M. Wang, P. Scambler, E. Custard, W. Reardon, S. Hassan, M. Muenke, P. Nixon, C. Papapetrou, R. M. Winter, Y. Edwards, K. Morrison, M. Barrow, M. P. Cordier-Alex, P. Correia, P. A. Galvin-Parton, S. Gaskill, K. J. Gaskin, S. Garcia-Minaur, R. Gereige, R. Hayward, T. Homfray, C. McKeown, V. Murday, H. Plauchu, N. Shannon, L. Spitz, and S. Lindsay]

Table 1 Clinical features and size of deletion of the 12 patients with 13q mono-

somy. NK: not known, m: months, y: years,WG: weeks of gestation, F: female, M: male; ARSA: Aberrant Right Subclavian Artery; DSCV: Double Superior Vena Cava. (1) Nasal bone hypoplasia from [Twelve new patients with 13q deletion syndrome: Genotypephenotype analyses in progress Author(s) Queacutèlin, Chloeacute; and Bendavid, Claude and Dubourg, Christegravèle and de la Rochebrochard, Ceacutèline and Lucas, Josette and Henry, Catherine and Jaillard, Sylvie and Loget, Philippe and Loeuillet, Laurence and Lacombe, Didier and Rival, Jean-Marie and David, Veacutèronique and Odent, Sylvie and Pasquier, Laurent]

Table 1 Summary of neuroimaging studies of four neurogenetic disorders from [Neural phenotypes of common and rare genetic variants Carrie E. Bearden, David C. Glahn, Agatha D. Lee, Ming-Chang Chiang, Theo G.M. van Erp, Tyrone D. Cannon, Allan L. Reiss, Arthur W. Toga, Paul M. Thompson]

TABLE I. List of the Clinical and Molecular Data from [Comprehensive Genetic Analysis of Relevant Four Genes in 49 Patients With Marfan Syndrome or Marfan-Related Phenotypes Author(s) Haruya Sakai, Remco Visser, Shiro Ikegawa, Etsuro Ito, Hironao Numabe, Yoriko Watanabe, Haruo Mikami, Tatsuro Kondoh, Hiroshi Kitoh, Ryusuke Sugiyama, Nobuhiko Okamoto, Tsutomu Ogata, Riccardo Fodde, Seiji Mizuno, Kyoko Takamura, Masayuki Egashira, Nozomu Sasaki, Sachiro Watanabe, Shigeru Nishimaki, Fumio Takada, Toshiro Nagai, Yasushi Okada, Yoshikazu Aoka, Kazushi Yasuda, Mitsuji Iwasa, Shigetoyo Kogaki, Naoki Harada, Takeshi Mizuguchi, and Naomichi Matsumoto]

TABLE II. LoeysDietz Syndrome Features in Patients With TGFBR2 or TGFBR1 Abnormality from [Comprehensive Genetic Analysis of Relevant Four Genes in 49 Patients With Marfan Syndrome or Marfan-Related Phenotypes Author(s) Haruya Sakai, Remco Visser, Shiro Ikegawa, Etsuro Ito, Hironao Numabe, Yoriko Watanabe, Haruo Mikami, Tatsuro Kondoh, Hiroshi Kitoh, Ryusuke Sugiyama, Nobuhiko Okamoto, Tsutomu Ogata, Riccardo Fodde, Seiji Mizuno, Kyoko Takamura, Masayuki Egashira, Nozomu Sasaki, Sachiro Watanabe, Shigeru Nishimaki, Fumio Takada,

Toshiro Nagai, Yasushi Okada, Yoshikazu Aoka, Kazushi Yasuda, Mitsuji Iwasa, Shigetoyo Kogaki, Naoki Harada, Takeshi Mizuguchi, and Naomichi Matsumoto]

TABLE III. Polymorphisms Found in This Study from [Comprehensive Genetic Analysis of Relevant Four Genes in 49 Patients With Marfan Syndrome or Marfan-Related Phenotypes Author(s) Haruya Sakai, Remco Visser, Shiro Ikegawa, Etsuro Ito, Hironao Numabe, Yoriko Watanabe, Haruo Mikami, Tatsuro Kondoh, Hiroshi Kitoh, Ryusuke Sugiyama, Nobuhiko Okamoto, Tsutomu Ogata, Riccardo Fodde, Seiji Mizuno, Kyoko Takamura, Masayuki Egashira, Nozomu Sasaki, Sachiro Watanabe, Shigeru Nishimaki, Fumio Takada, Toshiro Nagai, Yasushi Okada, Yoshikazu Aoka, Kazushi Yasuda, Mitsuji Iwasa, Shigetoyo Kogaki, Naoki Harada, Takeshi Mizuguchi, and Naomichi Matsumoto]

Table 1. PTEN mutations in CD and BZS from [Mutation spectrum and genotype-phenotype analyses in Cowden disease and BannayanZonana syndrome, two hamartoma syndromes with germline PTEN mutation Debbie J. Marsh, Valèrie Coulon, Kathryn L. Lunetta, Philippe Rocca-Serra, Patricia L. M. Dahia, Zimu Zheng, Danny Liaw, Stacey Caron, Bernadette Dubouè, Albert Y. Lin, Anne-Louise Richardson, Jean-Marie Bonnetblanc, Jean-Marie Bressieux, Agnès Cabarrot-Moreau, Agns Chompret, Liliane Demange, Rosalind A. Eeles, Alan M. Yahanda, Eric R. Fearon, Jean-Pierre Fricker, Robert J. Gorlin, Shirley V. Hodgson, Susan Huson, Didier Lacombe, Frèdèric LePrat, Sylvie Odent, Claude Toulouse, Olufunmilayo I. Olopade, Hagay Sobol, Sigrid Tishler, C. Geoffrey Woods, Bruce G. Robinson, H. Christian Weber, Ramon Parsons, Monica Peacocke, Michel Longy, and Charis Eng]

Table 4. Relationship between number of organ sites involved and specific PTEN mutation type or location in CD families from [Mutation spectrum and genotype-phenotype analyses in Cowden disease and BannayanZonana syndrome, two hamartoma syndromes with germline PTEN mutation Debbie J. Marsh, Valèrie Coulon, Kathryn L. Lunetta, Philippe Rocca-Serra, Patricia L. M. Dahia, Zimu Zheng, Danny Liaw, Stacey Caron, Bernadette Dubouè, Albert Y. Lin, Anne-Louise Richardson,

Jean-Marie Bonnetblanc, Jean-Marie Bressieux, Agnès Cabarrot-Moreau, Agns Chompret, Liliane Demange, Rosalind A. Eeles, Alan M. Yahanda, Eric R. Fearon, Jean-Pierre Fricker, Robert J. Gorlin, Shirley V. Hodgson, Susan Huson, Didier Lacombe, Frèdèric LePrat, Sylvie Odent, Claude Toulouse, Olufunmilayo I. Olopade, Hagay Sobol, Sigrid Tishler, C. Geoffrey Woods, Bruce G. Robinson, H. Christian Weber, Ramon Parsons, Monica Peacocke, Michel Longy, and Charis Eng]

Table 1. Sequences of MLPA Probes from [Loss-of-Function Mutations in Euchromatin Histone Methyl Transferase 1 (EHMT1) Cause the 9q34 Subtelomeric Deletion Syndrome Author(s) Tjitske Kleefstra, Han G. Brunner, Jeanne Amiel, Astrid R. Oudakker, Willy M. Nillesen, Alex Magee, David Geneviève, Valèrie Cormier-Daire, Hilde van Esch, Jean-Pierre Fryns, Ben C. J. Hamel, Erik A. Sistermans, Bert B. A. de Vries, and Hans van Bokhoven]

Table 2. Changes Found with Direct Sequencing of the EHMT1 Gene from [Loss-of-Function Mutations in Euchromatin Histone Methyl Transferase 1 (EHMT1) Cause the 9q34 Subtelomeric Deletion Syndrome Author(s) Tjitske Kleefstra, Han G. Brunner, Jeanne Amiel, Astrid R. Oudakker, Willy M. Nillesen, Alex Magee, David Geneviève, Valèrie Cormier-Daire, Hilde van Esch, Jean-Pierre Fryns, Ben C. J. Hamel, Erik A. Sistermans, Bert B. A. de Vries, and Hans van Bokhoven]

Table 3. Clinical Features of the Five Presented Patients Compared with the Phenotypic Characteristics of 9q Subtelomeric Deletion Syndrome from [Loss-of-Function Mutations in Euchromatin Histone Methyl Transferase 1 (EHMT1) Cause the 9q34 Subtelomeric Deletion Syndrome Author(s) Tjitske Kleefstra, Han G. Brunner, Jeanne Amiel, Astrid R. Oudakker, Willy M. Nillesen, Alex Magee, David Genevie'ève, Valèrie Cormier-Daire, Hilde van Esch, Jean-Pierre Fryns, Ben C. J. Hamel, Erik A. Sistermans, Bert B. A. de Vries, and Hans van Bokhoven]

Table 1 APC germline mutations identified in Greek FAP patients with supportive clinical data from [Mutational spectrum of APC and genotypephenotype correlations in Greek FAP patients Author(s) Florentia Fostira, Georgia Thodi, Raphael Sandalt-

zopoulos, George Fountzilas, Drakoulis Yannoukakos]

Table 1. Genes and primers for pyrosequencing from [Spatial, temporal and interindividual epigenetic variation of functionally important DNA methylation patterns Eberhard Schneider, Galyna Pliushch, Nady El Hajj, Danuta Galetzka, Alexander Puhl, Martin Schorsch, Katrin Frauenknecht, Thomas Riepert, Achim Tresch, Annette M. Muller, Wiltrud Coerdt, Ulrich Zechner and Thomas Haaf]

Table 1 Current nosographic status of NCL from [Correlations between genotype, ultrastructural morphology and clinical phenotype in the neuronal ceroid lipofusci Author(s) Sara E. Mole . Ruth E. Williams . Hans H. Goebel]

TABLE I. Main Clinical Features of the 42 JPS Patients from [Vessels Morphology in SMAD4 and BMPR1A-Related Juvenile Polyposis Author(s) Adriana Handra-Luca, Christel Condroyer, Cèline de Moncuit, Maryline Tepper, Jean-Francois Flèjou, Gilles Thomas, and Sylviane Olschwang]

TABLE II. Point Mutations Identified in the BMPR1A and SMAD4 Genes from [Vessels Morphology in SMAD4 and BMPR1A-Related Juvenile Polyposis Author(s) Adriana Handra-Luca, Christel Condroyer, Cèline de Moncuit, Maryline Tepper, Jean-Francois Flèjou, Gilles Thomas, and Sylviane Olschwang]

TABLE III. Morphological and Immunochemical Characterization of Juvenile Polyps According to the Mutated Gene from [Vessels Morphology in SMAD4 and BMPR1A-Related Juvenile Polyposis Author(s) Adriana Handra-Luca, Christel Condroyer, Cèline de Moncuit, Maryline Tepper, Jean-Francois Flèjou, Gilles Thomas, and Sylviane Olschwang]

["Correlation of mutations of the SH2D1A gene and Epstein-Barr virus infection with clinical phenotype and outcome in X-linked lymphoproliferative disease" Author(s) Janos Sumegi, Dali Huang, Arpad Lanyi, Jack D. Davis, Thomas A. Seemayer, Akihiko Maeda, George Klein, Marco Seri, Hiroshi Wakiguchi, David T. Purtilo and Thomas G. Gross]

Table 1. Sequence of the primers used to amplify exons, exon/intron junctions, and

presumed regulatory region of the SH201A from [Correlation of mutations of the SH2D1A gene and Epstein-Barr virus] Table 2. SH2D1A mutations in families with definitive diagnosis of XLP from [Correlation of mutations of the SH2D1A gene and Epstein-Barr virus infection with clinical phenotype and outcome in X-linked lymphoproliferative disease Author(s) Janos Sumegi, Dali Huang, Arpad Lanyi, Jack D. Davis, Thomas A. Seemayer, Akihiko Maeda, George Klein, Marco Seri, Hiroshi Wakiguchi, David T. Purtilo and Thomas G. Gross]

Table 4. Effect of EBV infection on clinical phenotype in XLP from [Correlation of mutations of the SH2D1A gene and Epstein-Barr virus infection with clinical phenotype and outcome in X-linked lymphoproliferative disease Author(s) Janos Sumegi, Dali Huang, Arpad Lanyi, Jack D. Davis, Thomas A. Seemayer, Akihiko Maeda, George Klein, Marco Seri, Hiroshi Wakiguchi, David T. Purtilo and Thomas G. Gross]

TABLE I. Mutations With Phenotypic Features from [Detection of 53 FBN1 Mutations (41 Novel and 12 Recurrent) and GenotypePhenotype Correlations in 113 Unrelated Probands Referred With Marfan Syndrome, or a Related Fibrillinopathy Author(s) C.L.S. Turner, H. Emery, A.L. Collins, R.J. Howarth, C.M. Yearwood, E. Cross, P.J. Duncan, D.J. Bunyan, J.F. Harvey, and N.C. Foulds]

TABLE I. Clinical Details of Patients of the Present Series With ZEB2 Mutations from [MowatWilson Syndrome: Facial Phenotype Changing With Age: Study of 19 Italian Patients and Review of the Literature Author(s) L. Garavelli, M. Zollino, P. Cerruti Mainardi, F. Gurrieri, F. Rivieri, F. Soli, R. Verri, E. Albertini, E. Favaron, M. Zignani, D. Orteschi, P. Bianchi, F. Faravelli, F. Forzano, M. Seri, A. Wischmeijer, D. Turchetti, E. Pompilii, M. Gnoli, G. Cocchi, L. Mazzanti, R. Bergamaschi, D. De Brasi, M.P. Sperandeo, F. Mari, V. Uliana, R. Mostardini, M. Cecconi, M Grasso, S. Sassi, G. Sebastio, A. Renieri, M. Silengo, S. Bernasconi, N. Wakamatsu, and G. Neri]

TABLE II. Clinical Findings in MWS Patients With ZEB2 Mutations from [MowatWilson Syndrome: Facial Phenotype Changing With Age: Study of 19 Italian Patients and Review of the Literature Author(s) L. Garavelli, M. Zollino, P. Cerruti Mainardi,

F. Gurrieri, F. Rivieri, F. Soli, R. Verri, E. Albertini, E. Favaron, M. Zignani, D. Orteschi, P. Bianchi, F. Faravelli, F. Forzano, M. Seri, A. Wischmeijer, D. Turchetti, E. Pompilii, M. Gnoli, G. Cocchi, L. Mazzanti, R. Bergamaschi, D. De Brasi, M.P. Sperandeo, F. Mari, V. Uliana, R. Mostardini, M. Cecconi, M Grasso, S. Sassi, G. Sebastio, A. Renieri, M. Silengo, S. Bernasconi, N. Wakamatsu, and G. Neri]

TABLE III. ZEB2 Mutations in the Present Series from [MowatWilson Syndrome: Facial Phenotype Changing With Age: Study of 19 Italian Patients and Review of the Literature Author(s) L. Garavelli, M. Zollino, P. Cerruti Mainardi, F. Gurrieri, F. Rivieri, F. Soli, R. Verri, E. Albertini, E. Favaron, M. Zignani, D. Orteschi, P. Bianchi, F. Faravelli, F. Forzano, M. Seri, A. Wischmeijer, D. Turchetti, E. Pompilii, M. Gnoli, G. Cocchi, L. Mazzanti, R. Bergamaschi, D. De Brasi, M.P. Sperandeo, F. Mari, V. Uliana, R. Mostardini, M. Cecconi, M Grasso, S. Sassi, G. Sebastio, A. Renieri, M. Silengo, S. Bernasconi, N. Wakamatsu, and G. Neri]

TABLE IV. ZEB2 Mutations in the Literature and in the Present Series from [MowatWilson Syndrome: Facial Phenotype Changing With Age: Study of 19 Italian Patients and Review of the Literature Author(s) L. Garavelli, M. Zollino, P. Cerruti Mainardi, F. Gurrieri, F. Rivieri, F. Soli, R. Verri, E. Albertini, E. Favaron, M. Zignani, D. Orteschi, P. Bianchi, F. Faravelli, F. Forzano, M. Seri, A. Wischmeijer, D. Turchetti, E. Pompilii, M. Gnoli, G. Cocchi, L. Mazzanti, R. Bergamaschi, D. De Brasi, M.P. Sperandeo, F. Mari, V. Uliana, R. Mostardini, M. Cecconi, M Grasso, S. Sassi, G. Sebastio, A. Renieri, M. Silengo, S. Bernasconi, N. Wakamatsu, and G. Neri]

TABLE I. Summary of the Clinical Findings in 23 Patients With NicolaidesBaraitser Syndrome from [NicolaidesBaraitser Syndrome: Delineation of the Phenotype Author(s) S'ergio B. Sousa, Omar A. Abdul-Rahman, Armand Bottani, Val'erie Cormier-Daire, Alan Fryer, Gabriele Gillessen-Kaesbach, Denise Horn, Dragana Josifova, Alma Kuechler, Melissa Lees, Kay MacDermot, Alex Magee, Fanny Morice-Picard, Elizabeth Rosser, Ajoy Sarkar, Nora Shannon, Irene Stolte-Dijkstra, Alain Verloes, Emma Wakeling, Louise Wilson, and Raoul C.M. Hennekam]

Table 1 Major clinical features of maternal UPD(14) syndrome from [Maternal uniparental disomy of chromosome 14 confined to an interstitial segment (14q23-14q24.2) Author(s) Rick A Martin, Darrin W Sabol, Peter K Rogan]

Table 2 Genotypic analysis from [Maternal uniparental disomy of chromosome 14 confined to an interstitial segment (14q23-14q24.2) Author(s) Rick A Martin, Darrin W Sabol, Peter K Rogan]

Table 1 Severity Scores, Age at Diagnosis, and Plasma Sterol Concentrations of Patients with SLOS from [Mutational Spectrum in the $\triangle$7-Sterol Reductase Gene and Genotype- Phenotype Correlation in 84 Patients with Smith-Lemli-Opitz Syndrome Author(s) M. Witsch-Baumgartner, B. U. Fitzky, M. Ogorelkova, H. G. Kraft, F. F. Moebius, H. Glossmann, U. Seedorf, G. Gillessen-Kaesbach, G. F. Hoffmann, P. Clayton, R. I. Kelley, and G. Utermann]

Table 2 Spectrum of SLOS-Causing Mutations in the DHCR7 Gene from [Mutational Spectrum in the $\triangle$7-Sterol Reductase Gene and Genotype- Phenotype Correlation in 84 Patients with Smith-Lemli-Opitz Syndrome Author(s) M. Witsch-Baumgartner, B. U. Fitzky, M. Ogorelkova, H. G. Kraft, F. F. Moebius, H. Glossmann, U. Seedorf, G. Gillessen-Kaesbach, G. F. Hoffmann, P. Clayton, R. I. Kelley, and G. Utermann]

Table 3 Frequent Mutations in the DHCR7 Gene of 84 Patients with SLOS from [Mutational Spectrum in the $\triangle$7-Sterol Reductase Gene and Genotype- Phenotype Correlation in 84 Patients with Smith-Lemli-Opitz Syndrome Author(s) M. Witsch-Baumgartner, B. U. Fitzky, M. Ogorelkova, H. G. Kraft, F. F. Moebius, H. Glossmann, U. Seedorf, G. Gillessen-Kaesbach, G. F. Hoffmann, P. Clayton, R. I. Kelley, and G. Utermann]

Table 4 Distribution of DHCR7 Genotype Classes, in Relation to Age at Diagnosis and Severity, in Patients with SLOS from [Mutational Spectrum in the $\triangle$7-Sterol Reductase Gene and Genotype- Phenotype Correlation in 84 Patients with Smith-Lemli-Opitz Syndrome Author(s) M. Witsch-Baumgartner, B. U. Fitzky, M. Ogorelkova, H. G. Kraft, F. F. Moebius, H. Glossmann, U. Seedorf, G. Gillessen-Kaesbach, G. F.

Hoffmann, P. Clayton, R. I. Kelley, and G. Utermann]

Table 5 Mutations in the DHCR7 Gene in Patients with SLOS Who Have a Severity Score ¿50 from [Mutational Spectrum in the △7-Sterol Reductase Gene and Genotype- Phenotype Correlation in 84 Patients with Smith-Lemli-Opitz Syndrome Author(s) M. Witsch-Baumgartner, B. U. Fitzky, M. Ogorelkova, H. G. Kraft, F. F. Moebius, H. Glossmann, U. Seedorf, G. Gillessen-Kaesbach, G. F. Hoffmann, P. Clayton, R. I. Kelley, and G. Utermann]

Table 1 PCR primers for amplification of exons 1, 2, and 3 from [Spectrum and distribution of MECP2 mutations in 64 Italian Rett syndrome girls: tentative genotype/phenotype correlation Author(s) L. Giunti, S. Pelagatti, V. Lazzerini, S. Guarducci, E. Lapi, S. Coviello, A. Cecconi, L. Ombroni, E. Andreucci, I. Sani, A. Brusaferri, A. Lasagni, G. Ricotti, B. Giometto, P. Nicolao, P. Gasparini, M. Granatiero, M.L. Giovannucci Uzielli]

Table 2 Sequencing primers from [Spectrum and distribution of MECP2 mutations in 64 Italian Rett syndrome girls: tentative genotype/phenotype correlation Author(s) L. Giunti, S. Pelagatti, V. Lazzerini, S. Guarducci, E. Lapi, S. Coviello, A. Cecconi, L. Ombroni, E. Andreucci, I. Sani, A. Brusaferri, A. Lasagni, G. Ricotti, B. Giometto, P. Nicolao, P. Gasparini, M. Granatiero, M.L. Giovannucci Uzielli]

Table 3 Identified mutations of MECP2 gene from [Spectrum and distribution of MECP2 mutations in 64 Italian Rett syndrome girls: tentative genotype/phenotype correlation Author(s) L. Giunti, S. Pelagatti, V. Lazzerini, S. Guarducci, E. Lapi, S. Coviello, A. Cecconi, L. Ombroni, E. Andreucci, I. Sani, A. Brusaferri, A. Lasagni, G. Ricotti, B. Giometto, P. Nicolao, P. Gasparini, M. Granatiero, M.L. Giovannucci Uzielli]

Table 4 The ten recurrent mutations characterized in a total of 48 unrelated patients (75.0syndrome girls: tentative genotype/phenotype correlation Author(s) L. Giunti, S. Pelagatti, V. Lazzerini, S. Guarducci, E. Lapi, S. Coviello, A. Cecconi, L. Ombroni, E. Andreucci, I. Sani, A. Brusaferri, A. Lasagni, G. Ricotti, B. Giometto, P. Nicolao,

P. Gasparini, M. Granatiero, M.L. Giovannucci Uzielli]

Table 5 Large deletions: seven large deletions were characterized within exon 3 beyond the TRD (between 1152 and 1202 bp, mostly overlapping), all causing the same premature terminate codon (PTC) at position 403 (P403X) from [Spectrum and distribution of MECP2 mutations in 64 Italian Rett syndrome girls: tentative genotype/phenotype correlation Author(s) L. Giunti, S. Pelagatti, V. Lazzerini, S. Guarducci, E. Lapi, S. Coviello, A. Cecconi, L. Ombroni, E. Andreucci, I. Sani, A. Brusaferri, A. Lasagni, G. Ricotti, B. Giometto, P. Nicolao, P. Gasparini, M. Granatiero, M.L. Giovannucci Uzielli]

Table 6 Tentative genotype/phenotype correlation: large deletions in the same area of exon 3 (see Table 5) from [Spectrum and distribution of MECP2 mutations in 64 Italian Rett syndrome girls: tentative genotype/phenotype correlation Author(s) L. Giunti, S. Pelagatti, V. Lazzerini, S. Guarducci, E. Lapi, S. Coviello, A. Cecconi, L. Ombroni, E. Andreucci, I. Sani, A. Brusaferri, A. Lasagni, G. Ricotti, B. Giometto, P. Nicolao, P. Gasparini, M. Granatiero, M.L. Giovannucci Uzielli]

Table 7 Tentative genotype/phenotype correlation in 12 Rett subjects with T158M mutation from [Spectrum and distribution of MECP2 mutations in 64 Italian Rett syndrome girls: tentative genotype/phenotype correlation Author(s) L. Giunti, S. Pelagatti, V. Lazzerini, S. Guarducci, E. Lapi, S. Coviello, A. Cecconi, L. Ombroni, E. Andreucci, I. Sani, A. Brusaferri, A. Lasagni, G. Ricotti, B. Giometto, P. Nicolao, P. Gasparini, M. Granatiero, M.L. Giovannucci Uzielli]

TABLE 1. Mutations Identified in Patients With Crigler-Najjar Syndrome Type 1 from [Genetic Lesions of Bilirubin Uridine-diphosphoglucuronate Glucuronosyltransferase (UGT1A1) Causing Crigler-Najjar and Gilbert Syndromes: Correlation of Genotype to Phenotype Author(s) Ajit Kadakol, Siddhartha S. Ghosh, Baljit S. Sappal, Girish Sharma, Jayanta Roy Chowdhury, and Namita Roy Chowdhury]

TABLE 2. Mutations of the UGT1A1 Coding Region That Result in Reduced Bilirubin Glucuronidation (Crigler-Najjar Syndrome Type 2 and Some Cases of Gilbert

Syndrome) from [Genetic Lesions of Bilirubin Uridine-diphosphoglucuronate Glucuronosyltransferase (UGT1A1) Causing Crigler-Najjar and Gilbert Syndromes: Correlation of Genotype to Phenotype Author(s) Ajit Kadakol, Siddhartha S. Ghosh, Baljit S. Sappal, Girish Sharma, Jayanta Roy Chowdhury, and Namita Roy Chowdhury]

Table 1 WSCR gene primer sequences from [Using case study comparisons to explore genotype-phenotype correlations in Williams-Beuren syndrome Author(s) A Karmiloff-Smith, J Grant, S Ewing, M J Carette, K Metcalfe, D Donnai, A P Read, M Tassabehji]

Table 1. Single-gene and obesity-related Mendelian disorders from [The Human Obesity Gene Map: The 2005 Update Author(s) Tuomo Rankinen, Aamir Zuberi, Yvon C. Chagnon, S. John Weisnagel, George Argyropoulos, Brandon Walts, Louis P'erusse, and Claude Bouchard]

Table 2. Murine models of obesity from [The Human Obesity Gene Map: The 2005 Update Author(s) Tuomo Rankinen, Aamir Zuberi, Yvon C. Chagnon, S. John Weisnagel, George Argyropoulos, Brandon Walts, Louis P'erusse, and Claude Bouchard]

Table 4. Evidence for association between markers of candidate genes with obesity-related phenotypes from [The Human Obesity Gene Map: The 2005 Update Author(s) Tuomo Rankinen, Aamir Zuberi, Yvon C. Chagnon, S. John Weisnagel, George Argyropoulos, Brandon Walts, Louis P'erusse, and Claude Bouchard]

Table 5. Evidence for the presence of linkage with obesity-related phenotypes from [The Human Obesity Gene Map: The 2005 Update Author(s) Tuomo Rankinen, Aamir Zuberi, Yvon C. Chagnon, S. John Weisnagel, George Argyropoulos, Brandon Walts, Louis P'erusse, and Claude Bouchard]

Table 3 Primer sequences for the ampliWcation of corresponding ATP7B exons and a promoter region, DHPLC gradient buVer, and applicative column temperatures for optimal resolution of PCR products from [Mutation analysis of the ATP7B gene and genotype/phenotype correlation in 227 patients with Wilson disease from Author(s)

Slavka Vrabelova, Ondrej Letocha, Marek Borsky, Libor Kozak]

Table 4 Distribution and frequency of 40 Wilson disease mutations in 400 mutant alleles of the ATP7B gene from [Mutation analysis of the ATP7B gene and genotype/phenotype correlation in 227 patients with Wilson disease from Author(s) Slavka Vrabelova, Ondrej Letocha, Marek Borsky, Libor Kozak]

Table 5 Characteristics of detected polymorphisms in ATP7B gene from [Mutation analysis of the ATP7B gene and genotype/phenotype correlation in 227 patients with Wilson disease from Author(s) Slavka Vrabelova, Ondrej Letocha, Marek Borsky, Libor Kozak]